ELUCIDATING THE ROLE OF PROTEIN COFACTORS IN RNA CATALYSIS USING RIBONUCLEASE P AS THE MODEL SYSTEM

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
The Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By

Hsin-Yue Tsai, B.S.

*****

The Ohio State University
2006

Dissertation Committee:
Dr. Venkat Gopalan, Advisor
Dr. Kathleen Boris-Lawrie
Dr. Tien-Hsien Chang
Dr. Mark P. Foster

Approved by
Advisor
Molecular, Cellular and Developmental Biology
ABSTRACT

Ribonuclease P (RNase P) catalyzes the 5′ maturation of tRNAs in all three domains of life and functions as a Mg\(^{2+}\)-dependent ribonucleoprotein (RNP) complex. It is composed of one RNA subunit, essential for catalysis, and a varying number of protein cofactors depending on the source. We have now used bacterial and archaeal RNase P to understand how proteins aid RNA catalysis.

Bacterial RNase P is composed of one catalytic RNA subunit and one protein cofactor, which is known to facilitate substrate binding and RNA catalysis. Although molecular modeling led to tertiary structure models of the RNA subunits, that were subsequently shown to be correct, and high-resolution studies established the structure of the protein subunit from bacterial RNase P, RNA-protein interactions in the holoenzyme were not established. Here, we have used a hydroxyl radical-mediated footprinting approach to generate this information which, together with results from other biochemical/biophysical studies, have furnished distance constraints for building
three-dimensional models of the bacterial RNase P holoenzyme in the absence or presence of its precursor tRNA substrate. The model reveals how the protein subunit facilitates RNA catalysis by directly interacting with both the ptRNA substrate and the catalytic core of the RNA subunit.

Unlike bacterial RNase P, both archaeal and eukaryal RNase P contain multiple protein subunits, whose roles are unclear largely due to the failure to reconstitute archaeal/eukaryal RNase P in vitro. Using recombinant subunits, we have now reconstituted functional RNase P from *Pyrococcus furiosus*, a thermophilic archaeon, and gained insights regarding its assembly pathway(s) and the contribution of its different protein subunits to RNA catalysis. The *Pfu* RNase P RNA is capable of multiple turnover catalysis with either of two pairs of protein subunits and becomes significantly more active, at lower magnesium concentrations, with addition of the remaining protein pair. These data support a central tenet of the RNA world hypothesis that the evolution of RNA enzymes to RNP complexes involved gradual recruitment of proteins to enhance biological function.
Collectively, these two studies highlight the common strategies employed by protein cofactors to enhance RNA catalysis (i.e., enhanced substrate binding and improved affinity for Mg$^{2+}$).
Dedicated to my mother, Shu-Lin Fu
ACKNOWLEDGMENTS

I express my deep and sincere gratitude to my advisor, Dr. Venkat Gopalan. Because of his intellectual guidance, unlimited support, and painstaking editing efforts, this dissertation has become possible.

I thank Drs. Mark Foster, Tien-Hsien Chang and Kathleen Boris-Lawrie, who have served in my dissertation committee and provided valuable feedback.

I am also grateful to some others who have contributed to this dissertation: Professor Eric Westhof, Institut de Biologie Moleculaire et Cellulaire-Centre National de la Recherche Scientifique (IBMC-CNRS), Strasbourg, FRANCE, and his research scientist, Benoit Masquida, for generating computer-aided tertiary models of the bacterial RNase P holoenzyme in the absence and presence of its substrate; Professors Sidney Altman, Yale University, and Edward Behrman, The Ohio State University, for valuable comments on our manuscripts.

I thank all the former and current members of the Gopalan Laboratory who have
helped me in one way or another. In particular, I am grateful to the following coworkers: Dr. Roopa Biswas for her mentoring; Dileep Pulukkunat for purifying native *Pyrococcus furiosus* (*Pfu*) RNase P holoenzyme and various valuable discussions; Walter Woznick for his help in cloning and purifying both RNA and protein subunits of *Pfu* RNase P; Cynthia Hatfield for discussion and technical support; and Lien Lai for valuable advice on cloning.

I am grateful to members of the Foster laboratory (OSU) for helpful discussions on archaeal RNase P and Kevin Fiala (Suo laboratory) for his helpful advice in setting up the Rapid Quench Flow instrument in our laboratory.

Finally, my special thanks to the American Heart Association for providing me two years of pre-doctoral fellowship support. I would also like to acknowledge the OSU MCDB graduate program, the Edward F. Hayes Graduate Research Forum, and the OSU College of Biological Sciences Herta Camerer Gross Fund for providing funding support either to attend meetings or for summer research.
VITA

3rd October 1975..........................Born - HsinChu, TAIWAN

1998.................................B.S. (Zoology), National Taiwan University, Taipei, TAIWAN

1998 - 1999 ..............................Research Assistant, Department of Zoology, National Academic Sinica, Taipei, TAIWAN

1999 - present ........................ Graduate Teaching and Research Associate, The Ohio State University, Columbus, OH

PUBLICATIONS


FIELDS OF STUDY

Major Field: Molecular, Cellular and Developmental Biology
TABLE OF CONTENTS

Abstract .....................................................................................................................................ii
Dedication ..................................................................................................................................v
Acknowledgments ...................................................................................................................vi
Vita .........................................................................................................................................viii
List of Tables ............................................................................................................................xiv
List of Figures ............................................................................................................................xv
List of Abbreviations ............................................................................................................xvii

Chapters:

1. Introduction ...........................................................................................................................1
   1.1 Prelude ..........................................................................................................................1
   1.2 RNase P ........................................................................................................................3
      1.2.1 Bacterial RNase P ................................................................................................4
          1.2.1.1 The RNA subunit of bacterial RNase P ...................................................5
          1.2.1.2 The protein subunit of bacterial RNase P ...............................................8
          1.2.1.3 Magnesium binding in bacterial RNase P .............................................10
          1.2.1.4 Substrate recognition of bacterial RNase P .........................................11
   1.2.2 Archaeal RNase P .............................................................................................13
      1.2.2.1 The RNA subunit of archaeal RNase P ...................................................14
      1.2.2.2 Protein subunits of archaeal/eukaryal RNase P ......................................15
      1.2.2.3 Structures of the protein subunits of archaeal RNase P .......................16
          1.2.2.3.1 Pop5 ................................................................................................16
          1.2.2.3.2 Rpp30 ..............................................................................................17
          1.2.2.3.3 Rpp21 ..............................................................................................18
          1.2.2.3.4 Rpp29 ..............................................................................................19
      1.2.2.4 Protein-protein interactions in archaeal RNase P ...................................20
1.3 Research objectives ....................................................................................................21
  1.3.1 Bacterial RNase P .............................................................................................21
  1.3.2 Archaeal RNase P .............................................................................................22

2. Molecular modeling of the three-dimensional structure of the bacterial RNase P holoenzyme .............................................................................................................................30
  2.1 Introduction ................................................................................................................30
  2.2 Material and methods .................................................................................................32
    2.2.1 Mutagenesis, purification, EPD-Fe modification and characterization of mutant derivatives of C5 protein ...............................................................................32
    2.2.2 Footprinting experiments ..................................................................................33
    2.2.3 Crosslinking assay and primer extension analysis ...........................................33
    2.2.4 Molecular modeling ..........................................................................................34
  2.3 Results .........................................................................................................................35
    2.3.1 Hydroxyl radical-mediated footprinting approach...........................................35
    2.3.2 EDTA-Fe modification and characterization of single Cys-substituted mutant derivatives of C5 protein ............................................................................................36
    2.3.3 Footprinting studies...........................................................................................37
    2.3.4 Non-covalent interactions in the ES complex ..................................................39
    2.3.5 Computer-aided molecular modeling ...............................................................40
  2.4 Discussion ...................................................................................................................43
    2.4.1 Directed hydroxyl radical-mediated footprinting strategy...............................43
    2.4.2 Location of the protein-binding site in the RNA subunit of bacterial RNase P .............................................................................................................................44
    2.4.3 Is the holoenzyme model consistent with previous data? .........................46
    2.4.4 Three-dimensional model of the ES complex ..................................................49
    2.4.5 Mechanism of action of bacterial RNase P and evolutionary considerations.51

3. Functional reconstitution and characterization of RNase P from Pyrococcus furiosus: insights into the transition of a catalytic RNA to an RNP complex ........................................66
  3.1 Introduction ................................................................................................................66
  3.2 Material and methods .................................................................................................68
    3.2.1 Cloning of the RNA and protein subunits of Pfu RNase P ...............................68
3.2.2 Generation of RPRs using \textit{in vitro} transcription .............................................. 69
3.2.3 Protein overexpression ............................................................................................. 70
3.2.4 Purification of protein subunits of \textit{Pfu} RNase P ................................................... 70
  3.2.4.1 \textit{Pfu} Pop5 ..................................................................................................... 70
  3.2.4.2 \textit{Pfu} Rpp29 ................................................................................................ 71
  3.2.4.3 \textit{Pfu} Rpp30 ................................................................................................ 71
  3.2.4.4 \textit{Pfu} Rpp21 ................................................................................................ 71
3.2.5 \textit{Pfu} RNase P reconstitution and assay .............................................................. 72
3.3 Results .......................................................................................................................... 73
  3.3.1 Nomenclature ....................................................................................................... 73
  3.3.2 Rationale .............................................................................................................. 74
  3.3.3 Purification of the recombinant subunits of \textit{Pfu} RNase P ............................... 75
  3.3.4 \textit{In vitro} reconstitution of \textit{Pfu} RNase P ......................................................... 75
  3.3.5 Reconstitution of partially and fully reconstituted \textit{Pfu} RNase P ..................... 77
3.4 Discussion .................................................................................................................... 79
  3.4.1 \textit{In vitro} reconstitution of \textit{Pfu} RNase P ........................................................... 79
  3.4.2 Hierarchical assembly of \textit{Pfu} RNase P .......................................................... 80
  3.4.3 Cooperativity during assembly of \textit{Pfu} RNase P .............................................. 82
  3.4.4 Role of protein cofactors/subunits in \textit{Pfu} RNase P catalysis ............................. 84
  3.4.5 Summary ........................................................................................................... 86
4 Discussion ....................................................................................................................... 95
  4.1 Summary ................................................................................................................... 95
  4.2 Molecular modeling of the three-dimensional structure of the bacterial RNase P
  holoenzyme .................................................................................................................... 96
  4.3 Functional reconstitution of the RNase P holoenzyme from \textit{Pyrococcus furiosus},
  a thermophilic archaean ............................................................................................. 98
  4.4 Conclusions ............................................................................................................. 99
5 Ongoing studies and future directions .......................................................................... 103
  5.1 Bacterial RNase P .................................................................................................... 103
  5.1.1 Two-site binding model to understand the interactions between bacterial RPR
  and Rpp ....................................................................................................................... 103
5.1.2 What is the contribution of the left-handed crossover in bacterial Rpp to RNase P catalysis? ........................................................................................................104

5.2 Archaeal RNase P.....................................................................................................105
   5.2.1 Role of protein subunits in archaeal RNase P catalysis.................................105
   5.2.2 Substrate recognition.......................................................................................106
   5.2.3 Molecular modeling of archaeal RNase P holoenzyme and ES complex....107

References.............................................................................................................................108
LIST OF TABLES

Page

2.1 Relative RNase P activities of various mutant derivatives of C5 protein (before and after EPD-Fe modification) and a summary of their footprinting results .................................54

2.2 Cα-PO4 distances in the model of *E. coli* RNase P ternary complex ..............................55

3.1 Characteristics of the protein subunits of *Pfu* RNase P .......................................................88

3.2 Oligonucleotide primers used for cloning the RNA and protein subunits of *Pfu* RNase P ........................................................................................................................................89
LIST OF FIGURES

Page

1.1 Role of RNase P in tRNA maturation.................................................................23

1.2 Classical representation of the 2D structure of the *E. coli* RNase P RNA and of the *B. subtilis* RNase P RNA .........................................................................................24

1.3 Comparison of the computer-aided three-dimensional models of bacterial RPRs and the tertiary structures recently established by X-ray crystallography .................................................26

1.4 Comparison of subunit composition of RNase P in all three domains of life ..........28

1.5 Tertiary structures of all four archaeal Rpps.......................................................29

2.1 Sites of mutagenesis to introduce Cys residues and subsequent tethering of EDTA-Fe56

2.2 OH-mediated cleavages in M1 RNA in the presence of unmodified and EDTA-Fe-modified derivatives of C5 protein.................................................................58

2.3 Summary of footprinting data ...........................................................................60

2.4 Crosslinking of the *E. coli* RNase P holoenzyme-substrate complex..............61
2.5 Model of the bacterial RNase P holoenzyme ................................................................. 62

2.6 Model of the bacterial RNase P holoenzyme-substrate complex .............................. 64

3.1 SDS-PAGE analysis illustrating the overexpression in *E. coli* and the homogeneity of the purified preparations of the four protein subunits of *Pfu* RNase P ........................................... 90

3.2 Functional reconstitution of *Pfu* RNase P with the RNA and protein subunits either alone or in various combinations ........................................................................................................ 91

3.3 Activities of partially and fully assembled *Pfu* RNase P ........................................ 92

3.4 Scheme of hierarchical assembly of *Pfu* RNase P ................................................. 94

4.1 Comparison of two different models for the interactions between a bacterial Rpp and the 5' leader sequence of a ptRNA substrate ................................................................. 101

4.2 Cladogram depicting hypothetical evolution of RNase P ...................................... 102
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>°C</td>
<td>Degree Celsius</td>
</tr>
<tr>
<td>3-D</td>
<td>Three-dimensional</td>
</tr>
<tr>
<td>A&lt;sub&gt;600&lt;/sub&gt;</td>
<td>Absorbance at 600 nm</td>
</tr>
<tr>
<td>Ala</td>
<td>Alanine</td>
</tr>
<tr>
<td>Arg</td>
<td>Arginine</td>
</tr>
<tr>
<td>AzP</td>
<td>Azidophenacyl</td>
</tr>
<tr>
<td>AzPBr</td>
<td>Azidophenacyl bromide</td>
</tr>
<tr>
<td>B. subtilis</td>
<td>Bacillus subtilis</td>
</tr>
<tr>
<td>C-domain</td>
<td>Catalytic domain</td>
</tr>
<tr>
<td>C-terminal</td>
<td>Carboxy-terminal</td>
</tr>
<tr>
<td>CD</td>
<td>Circular Dichroism</td>
</tr>
<tr>
<td>Cys</td>
<td>Cysteine</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>E. coli</td>
<td>Escherichia coli</td>
</tr>
<tr>
<td>EDTA-Fe</td>
<td>Ethylenediaminetetraacetic acid-iron</td>
</tr>
<tr>
<td>ES</td>
<td>Enzyme-substrate</td>
</tr>
<tr>
<td>h</td>
<td>Hour(s)</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>HEPES</td>
<td>N-(2-hydroxyethyl)piperazine-N’-(2-ethanesulfonic acid)</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>kDa</td>
<td>Kilodaltons</td>
</tr>
<tr>
<td>L</td>
<td>Liter</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani</td>
</tr>
<tr>
<td>M</td>
<td>Molar</td>
</tr>
<tr>
<td>min</td>
<td>Minute(s)</td>
</tr>
<tr>
<td>ml</td>
<td>Milliliter</td>
</tr>
<tr>
<td>mg</td>
<td>Milligram</td>
</tr>
<tr>
<td>mM</td>
<td>Millimolar</td>
</tr>
<tr>
<td>NAIM</td>
<td>Nucleotide analog interference mapping</td>
</tr>
<tr>
<td>ng</td>
<td>Nanograms</td>
</tr>
<tr>
<td>nm</td>
<td>Nanometer</td>
</tr>
<tr>
<td>nM</td>
<td>Nanomolar</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>N-terminal</td>
<td>Amino terminal</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>Pfu</td>
<td><em>Pyrococcus furiosus</em></td>
</tr>
<tr>
<td>ptRNA</td>
<td>Precursor transfer ribonucleic acid</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNase P</td>
<td>Ribonuclease P</td>
</tr>
<tr>
<td>RNP</td>
<td>Ribonucleoprotein</td>
</tr>
<tr>
<td>RNR motif</td>
<td>Arginine-Asparagine-Arginine motif</td>
</tr>
</tbody>
</table>
rpm  Rotations per minute

RRM  Ribonucleic acid-recognition motif

*S. aureus*  *Staphylococcus aureus*

S-domain  Specific domain

SDS-PAGE  Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

ssRNA  Single-stranded ribonucleic acid

*T. maritima*  *Thermotoga maritima*

tRNA  Transfer ribonucleic acid

UV  Ultraviolet

µg  Microgram(s)

µl  Microliter(s)

µm  Micrometer(s)
CHAPTER 1

INTRODUCTION

1.1 Prelude

The central role of RNA in various cellular processes such as translation, DNA replication, and RNA splicing was appreciated before the demonstration that RNAs could catalyze enzymatic reactions. The path-breaking findings two decades ago by Thomas Cech and Sidney Altman that the group I self-splicing intron and bacterial ribonuclease P (RNase P) RNA possess catalytic RNA moieties expanded the realm of biological catalysts (Guerrier-Takada et al, 1983; Kruger et al, 1982). Several RNA enzymes including the hairpin, hammerhead and Hepatitis delta virus (HDV) ribozymes have since been discovered (Chowrira & Burke, 1991; Keese & Symons, 1987; Hutchins et al, 1986; Wu et al, 1989). Collectively, these discoveries of RNA catalysts rekindled interest in the RNA world hypothesis that postulates RNA as the first macromolecule during the origin of life (Crick, 1968; Orgel, 1968; Gilbert, 1986). The dual ability to carry genetic information and promote catalysis of various reactions, possibly even self-copying, makes RNA a prime contender for being the first self-replicating molecule (Ekland & Bartel, 1996). Even though no organism operating solely with RNA catalysts has been discovered, several in
vitro evolution studies have demonstrated the potential of RNA to perform the basic enzymatic reactions of life: nucleotide synthesis, RNA polymerization, aminoacylation and peptide formation, and aldol condensation (Fusz et al, 2005; Johnston et al, 2001; Lee et al, 2000; Unrau & Bartel, 1998; Zhang & Cech, 1997). There is thus a growing body of evidence attesting to RNA’s potential to serve as a functional polymer essential for early life.

In extant life forms, protein enzymes with higher catalytic efficiency and versatility perform nearly all the cellular enzymatic reactions. However, many universal macromolecular catalysts still rely on RNA, perhaps reflecting a primordial stage dominated by RNA enzymes. For example, high-resolution structural studies of the ribosome reveal that there is a complete absence of proteins around the active site suggesting that the functional groups necessary for catalyzing peptide bond formation might reside solely in the rRNAs (Ban et al, 2000; Wimberly et al, 2000). Similarly, two snRNAs in the spliceosomal ribonucleoprotein (RNP) complex, which is made up of five snRNAs and several hundred proteins, can catalyze, in the absence of any spliceosomal protein components, the formation of a lariat structure, akin to the product of the first step of mRNA splicing. (Valadkhan & Manley, 2001). Moreover, an essential RNA subunit is present in RNase P, an RNP enzyme that catalyzes tRNA maturation in all three domains of life (Guerrier-Takada et al, 1983).

Despite the central role of RNA in the functioning of the ribosome, spliceosome and RNase P, these macromolecular machines function as ribonucleoproteins i.e., RNA + protein complexes. If indeed life evolved using RNA as the first macromolecule, why were protein cofactors recruited to participate in RNA-mediated catalysis in RNP complexes?
Were protein cofactors gradually mobilized to help RNA catalysts cope with new challenges imposed by the unique physiological situations encountered in the three domains of life? Despite considerable discussion on the hypothetical transition from an RNA to the extant RNA/protein world, the lack of molecular fossils has made the RNA world hypothesis difficult to test. RNase P, due to its increasing protein:RNA ratio as we proceed along the evolutionary path from bacteria to archaea to eukarya, is an evolutionary paradigm ideal for understanding such RNA to RNP transitions and forms the focus of this thesis.

1.2 RNase P

RNase P is an endoribonuclease that catalyzes 5’ maturation of precursor tRNAs (ptRNA) in all three domains of life: bacteria, archaea and eukarya (Fig. 1.1; Guerrier-Takada et al, 1983; Kirsebom & Vioque, 1995; Frank & Pace, 1998; Hall & Brown, 2001; Jarrous & Altman, 2001; Gopalan et al, 2002; Xiao et al, 2002; Hartmann & Hartmann, 2003; Hsieh et al, 2004). RNase P is a metalloenzyme since it requires magnesium for catalysis. It is an RNP complex consisting of both RNA and protein subunit(s). The RNA and protein subunits are abbreviated as RPR and Rpp, respectively. Even though RNase P in all domains of life contains one RPR, the number of Rpps varies: one in bacteria, at least four in archaea and up to ten in eukarya (Chamberlain et al, 1998; Hall & Brown, 2002; Jarrous, 2002; Kole & Altman, 1981).
1.2.1 Bacterial RNase P

Bacterial RNase P is a heterodimer consisting of one catalytic RNA subunit (size: ~400 nts; 120 kDa) and one protein cofactor (size: ~120 amino acid residues; 14 kDa). Bacterial RPR was the first true multiple-turnover RNA catalyst to be discovered; it catalyzes precursor tRNA maturation in trans at non-physiological ionic conditions \textit{in vitro} (Guerrier-Takada et al, 1983). However, both the RNA and protein subunits are absolutely essential for cell viability \textit{in vivo} (Schedl & Primakoff, 1973; Kirsebom et al, 1988).

While precursor tRNAs are the primary substrates of bacterial RNase P, the precursors to 4.5S RNA, tmRNA, some viral RNAs, C4 antisense RNA from bacteriophage P1 and P7, and the intergenic \textit{lac}YA region of the \textit{lac} operon mRNA are also cleaved by this enzyme \textit{in vitro} (Hartmann et al, 1995; Komine et al, 1994; Li & Altman, 2004; Mans et al, 1990; Peck-Miller & Altman, 1991). The physiological relevance of bacterial RNase P-mediated processing of some of these non-ptRNA substrates remains unclear.

RNase P requires Mg\textsuperscript{2+} for catalysis. The proposed catalytic mechanism of bacterial RNase P involves an S\textsubscript{N}2 in-line attack by a metal (preferably Mg\textsuperscript{2+})-bound hydroxide nucleophile positioned proximal to the scissile phosphodiester bond (i.e. the phosphodiester linkage immediately 5’ to the first nucleotide in mature tRNAs). A Mg\textsuperscript{2+} ion is also believed to stabilize the developing negative charge on the leaving group (Warnecke et al, 1996). In addition to direct roles in catalysis, Mg\textsuperscript{2+} has also been found to stabilize the tertiary structure of the RNA subunit and enhance the affinity of the RNA subunit for the ptRNA substrate (Pan, 1995; Persson et al, 2003).
1.2.1.1 The RNA subunit of bacterial RNase P

The secondary structure of bacterial RPR was first deduced by extensive comparative phylogenetic analysis and provided the rationale for classifying bacterial RPRs into two groups: type A (exemplified by *Escherichia coli*) and B (exemplified by *Bacillus subtilis*) (Brown, 1997). While several secondary structure elements are different between the type A and B versions, all bacterial RPRs share a minimal core of conserved sequences and secondary structures (Haas et al, 1994). As shown in Figure 1.2, secondary structures from P1 to P5 are conserved in both type A and B bacterial RPRs with a majority of conserved nucleotides clustered in the P4 region. The cruciform structure (consisting of P7 to P11) and J11/12 are also common among all bacterial RPRs with a majority of conserved nucleotides in J11/12.

The expectation that large RNAs might possess independent folding modules like their protein counterparts led to a search for such modules. Indeed, deletion studies revealed that the RNA subunit of bacterial RNase P can be divided into two domains: the catalytic (C) domain and substrate specificity (S) domain (Figs. 1.2C and D; Guerrier-Takada & Altman, 1992; Pan, 1995). The C-domain contains the minimal elements to perform catalytic activity. The minimal elements include important and highly conserved nucleotides in the RPR necessary for chemical cleavage and Watson-Crick base pairing with both the 5′ and 3′-ends of the ptRNA substrate such that the scissile bond is positioned optimally in the active site (Kirsebom et al, 1994; Oh & Pace, 1994; Heide et al, 1999; Zahler et al, 2003). The S-domain specifically interacts with the T-stem loop structure in ptRNAs (Pan et al, 1995). Interestingly, while the secondary structure predictions of the type A and B bacterial RPRs are quite different, recently established crystal structures
reveal a fairly similar tertiary fold including a concave substrate-binding cleft (Krasilnikov et al, 2003).

The low-resolution tertiary structure models for both type A and B bacterial RPRs were generated based on biochemical and biophysical studies several years before their crystal structures were solved (Figs. 1.3 A and B; Chen et al, 1998 and Massire et al, 1998). In the C-domain, the models for both type A and B bacterial RPRs are composed of helices, some of which stack on one another (Fig 1.3). The first stack includes P2 and P3 with an additional P19 helix only in type B RPR. The second stack, which is located in the central region of bacterial RPR, contains P1, P4, and P5. This stack is further elongated in type B RPR due to addition of P5.1. The third structural element is P15. Even though P15 is located in a similar position in both type A and B RPRs, it is further co-stacked with P16, P17 and P6 to form a long contiguous helix in type A RPR. The fourth motif, although positioned in spatially equivalent regions in type A and B RPRs, consists either of the P15.1/P15.2 stack in type B RPR or just P18 in type A RPR.

In the S-domain, two stacking structures were also proposed in bacterial RPRs from both types. The first stack is formed by P10/P11 that is common to both types A and B RPRs. The P13/P14 stack in type A RPR is believed to be analogous to the P10.1a/ P10.1 stack in type B RPR. Furthermore, the cruciform structure, which is thought to be the interface between C- and S-domains was proposed to be formed by two stacks with one formed by P8/P9 and the other joining P7 with P10/P11 in the S-domain (Massire et al, 1998).

Remarkably, the predicted models have no gross differences with the structures established recently by x-ray crystallography (Figs. 1.3C and D; Kazantsev et al, 2005;
Torres-Larios et al, 2005). The crystal structures of bacterial RPR also confirm biochemical observations that both S- and C- domains can be folded independently and reconstituted to generate a functional RNA catalyst (Loria & Pan, 1996). The two domains are linked by the P8/P9 stack. In type A bacterial RPR, the C-domain interacts with the P8/P9 stack (S-domain) through the P1-L9 and L18-P8 interactions (Fig. 1.3C). All the tertiary interactions thought to stabilize the overall fold of type A bacterial RPR were intra-molecular contacts predicted by co-variation analysis (Fig. 1.3C). For example, the long-range interaction between P1 and L9 was established by observing a correlation between the presence of a GNRA tetraloop sequence in L9 and a G/C at the eighth base pair of stem P1 (bp 4/370 in *E. coli* numbering; see Fig. 1.2A). Such a correlation was taken to suggest the conservation of a tertiary contact between the L9 tetraloop and the shallow groove of the P1 helix, a widely observed long-range structural motif (Massire et al, 1997).

The crystal structure of the C-domain displays features already predicted by the computer-aided model of type A bacterial RPRs: (1) a long stem formed by the co-axially arranged P2 and P3 helices; (2) a second long stem formed by the co-axially stacked P1, P4 and P5 helices; (3) a stem that folds on itself (P15/P17) and whose end forms a pseudoknot with P6; (4) a third stem, P18, which runs almost perpendicular to the other three helices (Fig. 1.3C). In the crystal structure of type B bacterial RPR (Kazantsev et al, 2005), one unique feature is that all the coaxially stacked helical subdomains form an unusually flat-faced structure. Moreover, the crystal structure of type B RPR reveals a complicated structure formed by two inter-helix base pairing interactions: P15.1/L5.1 and P4/L8 which were not modeled in the computer-aided models built earlier (Fig. 1.3D).
The agreement in the global orientation of all the helices between the predicted structural models and crystal structures of both type A and B RPRs lends credibility to various conclusions about structure-function relationships in bacterial RNase P that were drawn from the model.

1.2.1.2 The protein subunit of bacterial RNase P

The tertiary structure of bacterial RPP from *B. subtilis*, *Thermotoga maritima*, and *Staphylococcus aureus* has been determined using X-ray crystallography and NMR spectroscopy (Kazantsev et al, 2005; Spitzfaden et al, 2000; Stams et al, 1998). Despite the low sequence homology among the protein subunits of bacterial RNase P, all bacterial Rpps exhibit an identical tertiary fold. The structure of bacterial Rpps adopt an α-β sandwich fold comprising a central, four-stranded β-sheet surrounded by three α-helices. One face of the β-sheet packs with helices α2 and α3 to form the hydrophobic core while the other face is part of a cleft made possible by the curvature of the β-sheet and the packing of helix α1 (Fig. 4B). The α2 helix has the most conserved sequences among all bacterial Rpps including the RNR motif (AHxxRNRxxKRLxR, where x is any amino acid residue) (Fig. 4A). Another unique structural feature in bacterial Rpps is the left-handed crossover, which connects β3 to α2 to β4. This special structure is present in only few proteins and usually plays an important role in catalysis. The left-handed crossover was first identified in subtilisin (Wright et al, 1969) and has since been discovered to be present in acetycholine esterase (Sussman et al, 1991), steroid dehydrogenase (Ghosh et al, 1991), and L-asparaginase (Miller et al, 1993).

Even though the protein subunit of bacterial RNase P does not have any catalytic
activity, studies using a temperature-sensitive Rpp mutant in *E. coli* RNase P revealed its essentiality for cell survival (Schedl & Primakoff, 1973; Kirsebom et al, 1988). Due to the preponderance of positively charged residues in all bacterial Rpps and their ability increase the substrate binding affinity, it is believed that bacterial Rpp contributes to electrostatic neutralization between the polyanionic enzyme and substrate RNAs by providing a local pool of counter-ions that facilitates substrate binding without interfering with rapid product release (Reich et al, 1988). This simplistic notion is belied by the inability of polyamines to replace bacterial Rpp. Detailed mechanistic studies have now uncovered different roles of bacterial Rpp in RNase P catalysis. First, bacterial RPP enhances the binding affinity between bacterial RPR and ptRNA by binding the leader sequence in ptRNA. Crosslinking studies have shown that the leader sequence of ptRNA, specifically nucleotides -4 to -8 from the cleavage site, interact directly with the central cleft region of the bacterial Rpp in the ES complex (Niranjanakumari et al,1998). Consistent with this observation, pre-steady state kinetic studies revealed that shortening the leader sequence to only 2 or 3 nucleotides decreases the binding affinity of the RNase P holoenzyme for ptRNAs without influencing catalysis (Crary et al, 1998). Second, the addition of *B. subtilis* Rpp increases the rate of ptRNA cleavage by 10-fold compared to the reaction catalyzed by *B. subtilis* RPR alone (Kurz et al, 1998). Third, bacterial Rpp decreases the requirement for magnesium ions. The decrease was demonstrated by comparing the binding affinity of *B. subtilis* RPR and RNase P holoenzyme for Mg$^{2+}$ (Kurz & Fierke, 2002). This appears to be a common theme among metal-dependent-RNP enzymes since a similar protein-mediated decrease in Mg$^{2+}$ requirement has also been reported in mitochondrial self-splicing group I intron in the mitochondrial cytochrome b (COB) mRNA from *Aspergillus nidulans* (Solem et al, 2002).
Fourth, *E. coli* Rpp promotes RNA folding. While *E. coli* RPR with mutations in P8 miscleaves ptRNA substrates in the absence of its protein cofactor, addition of *E. coli* Rpp abolishes the tendency for miscleavage (Guerrier-Takada et al, 1989). Recently, Buck et al. (2005) also used gel-shift and temperature-gradient gel electrophoresis assays to demonstrate that *E. coli* RPP stabilizes the global tertiary structure of *E. coli* RPR against chemical and temperature denaturation (Buck et al, 2005).

### 1.2.1.3 Magnesium binding in bacterial RNase P

Magnesium is known to be important for RNase P catalysis due to its role in RNA folding, substrate binding and catalysis (Brannvall et al, 2001; Guerrier-Takada et al, 1986; Persson et al, 2003). While the binding of over 100 Mg$^{2+}$ ions independently and non-specifically helps neutralize the highly negatively charged bacterial RPR, only a small number (<10) of Mg$^{2+}$ ions bind specifically to facilitate RPR folding and substrate binding (Beebe et al, 1996). By measuring the cooperative binding between magnesium and bacterial RPR using a Hill plot, it was concluded that at least three Mg$^{2+}$ ions are required for RNase P catalysis (Smith & Pace, 1993).

Nucleotides in bacterial RPR or ptRNA important for Mg$^{2+}$ coordination have been analyzed by using nucleotide analog interference mapping (NAIM). Mg$^{2+}$, which binds specifically to bacterial RPR, is believed to interact with the RNA molecule by coordinating the oxygen atoms in the phosphate backbone. If substitution of an oxygen atom by a sulfur atom in the phosphate backbone results in a loss of activity, it indicates the possibility that the sites of substrates (bridging or non-bridging positions) are involved in Mg$^{2+}$ binding. Confirmatory evidence is obtained when a thiophilic metal ion (such as
Cd\textsuperscript{2+} or Mn\textsuperscript{2+} rescues the loss of activity observed with the phosphorothiate-containing RNA. Such NAIM studies have been performed on both the RPR and ptRNA substrates.

NAIM studies revealed that three Rp oxygen atoms located in the P4 region of bacterial RPR (A67, G68, A352) are vital for catalysis (Christian et al, 2000). Moreover, by exploiting a ptRNA binding assay, several Rp oxygen atoms were also identified as critical sites for coordinating Mg\textsuperscript{2+} between bacterial RPR and ptRNA (Warnecke et al, 1999; Pfeiffer et al, 2000). The nucleotides in the P15 region of bacterial RPR that were originally identified as important for substrate binding using NAIM have now been confirmed as a metal-binding pocket using the Pb\textsuperscript{2+}-induced cleavage method (Ciesiolka et al, 1994; Zito et al, 1993).

1.2.1.4 Substrate recognition by bacterial RNase P

The ptRNAs for all 64 codons display very low sequence homology but similar tertiary structures. Therefore, RNase P is thought to recognize its ptRNA substrates by their secondary and tertiary structural elements instead of primary sequences. However, studies using bacterial RNase P have shown that the limited but highly conserved sequence elements in ptRNA do contribute to substrate recognition. First, the two Cs in the CCA sequence present at the 3’ termini of most ptRNAs base pair with G292 and G293 (in L15) in bacterial RPR (Note: G292 and G293 are based on \textit{E. coli} RPR numbering) (Kirsebom & Svard, 1994). This base-pairing interaction anchors ptRNA to bacterial RPR. Second, sequence analysis of ptRNAs reveals that U and G are preferentially present at the -1 (the location immediately upstream of the RNase P cleavage site) and +1 position, respectively. Mutagenesis of the -1 position of a ptRNA and compensatory base changes at the
universally conserved A248 (J5/15) in *E. coli* RPR provided evidence of base pairing between these regions in the ES complex (Zahler et al, 2003). Lastly, specific 2′-OH groups in the T-loop (e.g., position 62) are contacted by conserved adenosines (e.g., A230 in *B. subtilis*) in the P10/P11 region of the S-domain of *B. subtilis* RPR; this was established by using a circular permutation analysis coupled with dephosphorylation strategy (Pan et al, 1995).

Other experiments examining the minimal structural elements required for bacterial RNase P recognition have established that the length of the leader sequence needs to be at least three nucleotides to promote efficient substrate binding. Both crosslinking and pre-steady state kinetic studies have shown that the leader sequence of ptRNA contributes tremendously to the binding affinity of the bacterial RNase P holoenzyme for ptRNA, presumably due to docking of leader sequence in the central cleft region of bacterial RPP (Niranjanakumari et al, 1998). The length of the acceptor stem, generated from stacking of the acceptor stem on the T-stem, also affects cleavage-site selection. Kirsebom and coworkers have shown that extending the acceptor stem (to greater than 12 bps) results in miscleavage of ptRNAs by bacterial RPR (Kufel & Kirsebom, 1996).

An Rp-phosphorothioate substitution at the -1 position in ptRNA resulted in a 10,000-fold decrease in catalytic efficiency. Rescue by both Mn$^{2+}$ and Cd$^{2+}$ suggests that the Rp oxygen atom in the -1 position in ptRNA is also involved in Mg$^{2+}$ coordination (Warnecke et al, 1999; Pfeiffer et al, 2000).

By integrating the information on the important sequence and structural elements used by bacterial RNase P for ptRNA recognition, a minimal substrate has been generated. This minimal substrate consists of two complementary RNA oligos that are 10 and 11
nucleotides long, respectively. Each RNA oligo has seven continuous nucleotides complementary to each other. The first three nucleotides at 5′ end of the 10-mer RNA oligo serve as the leader sequence of the substrate. The last four nucleotides, ACCA, at the 3′ terminus of the 11-mer RNA oligo mimics the 3′ terminus of all the bacterial ptRNAs. This model substrate was shown to be cleaved by bacterial RNase P about two orders of magnitude slower than wild-type ptRNA (Hansen et al, 2001).

1.2.2 Archaeal RNase P

Even though RNase P is a universal enzyme, the composition of RNase P is quite different in the three domains of life. RNase P is an RNP complex with one RNA subunit and varying number of protein components: one, at least four and up to ten protein subunits in bacterial, archael and eukaryal RNase P, respectively (Fig. 1.4). The predicted secondary structures of archaeal RPRs are similar to the bacterial counterparts; however, only some archaeal RPRs are catalytically active in vitro under conditions of high ionic strength (300 mM MgCl₂ and 4 M ammonium acetate) and in the absence of their protein subunits (Pannucci et al, 1999). The archaeal Rpps were first identified by examining the Methanothermobacter thermoautotrophicus genome for the presence of sequences homologous to yeast Rpps. Four putative archael Rpps homologous to Pop5, Rpr2, Pop4 and Rpp1 in the yeast RNase P were identified; these are also homologous to Pop5, Rpp21, Rpp29 and Rpp30 of human Rpps (Hall & Brown, 2002). These four proteins have further been experimentally verified as being associated with the archael RNase P holoenzyme by virtue of their co-elution with RNase P activity in column chromatographic elution profiles of native enzyme preparations (Hall & Brown, 2002). Moreover, the success of
reconstituting *Pyrococcus horikoshii* (*Pho*) RNase P using *Pho* RPR and four protein subunits, which are homologous to the protein subunits identified in *Mth*, further confirmed that these four protein subunits are part of archaeal RNase P (Kouzuma et al, 2003). Note that although *Pho* RNase P has been reconstituted successfully, the enzymatic activity of the reconstituted enzyme is both lower than that of native archaeal RNase P and under single-turnover conditions.

1.2.2.1 The RNA subunit of archaeal RNase P

Archaeal RPRs have also been categorized into two groups, types A and M (the latter primarily from *Methanococci*; Harris et al. 2001). Archaeal type A RPRs, exemplified by *Pyrococcus furiosus* (*Pfu*), are strikingly similar to the bacterial type A RPRs. The presence of P18 and P13/P14 in bacterial type A RPRs distinguish them from the archaeal type A RPR (Fig. 1.4). The P18 helix in bacterial type A RPR is highly conserved both in the number of base-pairs (8 bp, flanked by an additional two bases on each side) and also its loop sequence (GNRA; L18). L18 has been shown to interact with P8 to stabilize the cruciform structure. The absence of P8/L18 interaction was postulated to be the basis for the weaker activity and high salt requirement in archaeal RPR compared to bacterial RPR. However, the addition of a bacterial RPR P18 element to the *Methanobacterium thermoautotrophicum* (*Mth*) RPR neither improved the catalytic activity nor decreased the requirement for either monovalent or divalent ions. The absence of P13/P14 region might also result in destabilizing the tertiary structure of archaeal RPR, since the P13/14 structure in bacterial type A RPR provides a tertiary interaction between L14 and P8 within the S-domain (Massire et al, 1998; Torres-Larios et al, 2005).
Archaeal type M RPRs, exemplified by *Methanococcus jannaschii* (*Mja*), differ from archaeal type A RPR in their inability to show catalytic activity in the absence of protein subunits (Pannucci et al., 1999). The major structural differences between archaeal types A and M RPR are (i) absence of P8, and (ii) everything distal to P15, including L15 (Harris et al. 2001). While P8 in bacterial RPR is involved in substrate T-loop recognition (Pan et al., 1995), and L15, a Mg$^{2+}$-binding pocket, is known to base pair with 3’-CCA of ptRNA substrates (Heide et al., 1999; Kirsebom & Svärd, 1994; Kufel & Kirsebom, 1998; Oh & Pace, 1994), there are no obvious additional structural domains in the type M RPR to compensate for the absence of these elements.

### 1.2.2.2 Protein subunits of archaeal/eukaryal RNase P

Four archaeal Rpps have been identified and named as Pop5, Rpp21, Rpp29, and Rpp30 in accord with the human RNase P nomenclature. Without a robust *in vitro* reconstitution established for either archaeal or eukaryal RNase P, there is only limited information on the function of their protein subunits in catalysis. Since it is known that all yeast Rpps are essential for viability (Chamberlain et al., 1998), elucidation of structure-function relationships in these Rpps is essential to understand RNase P catalysis. Recently, there have been remarkable advances in establishing the structures of all four archaeal Rpps. These structures are described below.
1.2.2.3 Structures of the protein subunits of archaeal RNase P

1.2.2.3.1 Pop5

The tertiary structure of Pop5 from both *Pho* and *Pfu* have been determined using X-ray crystallography (Fig. 1.5 A; Wilson et al, 2006; Kawano et al, 2006). *Pfu* Pop5 is the only protein subunit in either archaeal or eukaryal RNase P that has a tertiary fold similar to the sole protein subunit from bacterial RNase P. Interestingly, the different secondary structure arrangements in *Pfu* Pop5 and bacterial Rpps (βααβααβαα in *Pfu* Pop5 versus αβββαβαα in the bacterial RPP) suggest that the two proteins might have different evolutionary origins. *Pfu* Pop5 adopts an α-β sandwich fold comprised of a central, four-stranded antiparallel β-sheet surrounded by four α-helices. While one face of the β-sheet packs with helices α₁, α₂, and α₃ forming the hydrophobic core, the other face is more exposed to solvent. The solvent-exposed side of the β-sheet contains three apolar side chains and loosely packs with six apolar side chains from helix α₄. Three major differences have been observed between *Pfu* Pop5 and bacterial Rpps. First, the central cleft in bacterial Rpp (Kazantsev et al, 2003; Spitzfaden et al., 2000; Stams et al., 1998), which is thought to directly interact with the leader sequence of ptRNA, is comprised of a four-stranded β-sheet and one α-helix (α₁) in bacterial Rpp. Helix α₄ in *Pfu* Pop5, which is analogous to α₁ in bacterial Rpp, does not directly pack with the β-sheet to form a central cleft (Wilson et al, 2006). Nevertheless, several conserved hydrophobic residues in the β-sheet of *Pfu* Pop5 might be expected to play a role in binding the leader sequence of ptRNA (akin to bacterial Rpp). Second, the RNR motif in α₂ of bacterial Rpps, which is highly conserved among all the bacterial Rpps and is essential for facilitating RPR catalysis,
is absent in *Pfu* Pop5. Third, the unique left-hand crossover structure in bacterial RPP is also not observed in *Pfu* Pop5.

The structure of *Pfu* Pop5 is homologous to one of the most abundant protein domains in eukarya, the RNA recognition motif (RRM), also known as the ribonucleoprotein domain (RNP) or the RNA-binding domain (RBD). RRM domain-containing proteins perform important roles in post-transcriptional events, including RNA processing, splicing and editing (Maris et al, 2005).

### 1.2.2.3.2 Rpp30

The tertiary structure of Rpp30 from *Pyrococcus horikoshii* (*Pho*) was also solved using X-ray crystallography (Fig. 1.5B; Takagi et al, 2004). The structure of Rpp30 is an $\alpha/\beta$ barrel structure, similar to the TIM barrel structure (Reardon & Farber, 1995). It contains 10 $\alpha$-helices and seven $\beta$-strands and forms an oblate ellipsoid. Rpp30 is not a conventional TIM barrel, which contains eight $\alpha$-helices and eight $\beta$-strands. Note that $\alpha_9$ in Rpp30 replaces $\beta_8$ in TIM barrel structure. Helix $\alpha_{10}$ in Rpp30 is situated at the bottom of the barrel pore, acting as a lid to the barrel pore. Helices $\alpha_{6-8}$ together with $\beta_{6,7}$ form a hydrophobic cluster at one edge of the molecule. Rpp30 (pI=10) has most of its positively charged residues located on one side of the cluster of $\alpha$-helices ($\alpha_{3-5}$). Rpp30 resembles cytosine deaminase from *E. coli* in both structure and sequence suggesting that both of them might have arisen from a common ancestor. Among the RNA binding proteins, Rpp30 is also similar to tRNA-guanine transglycosylase (TGT), which has a TIM barrel structure with tRNA binding sites located between $\alpha_7/\beta_7$ and $\alpha_8/\beta_8$. It is conceivable that Rpp30 might have similar binding sites for ptRNA (Nishimura, 1983).
1.2.2.3.3 Rpp21

The crystal structure of *Pho* Rpp21 has been solved (Fig. 1.5C; Kakuta et al, 2005). *Pho* Rpp21 is a Zn$^{2+}$-binding protein containing a zinc-ribbon domain. The presence of Zn$^{2+}$ is important for reconstituted *Pho* RNase P activity. The structure of Rpp21 can be divided into the N-terminal, central, and C-terminal domains. The N-terminal domain of Rpp21 is composed of two long α helices that interact through hydrophobic residues. The central domain is mainly an unstructured loop. The C-terminal region of Rpp21 contains a zinc-ribbon domain. It contains a three-stranded antiparallel β-sheet with a single zinc ion that is coordinated by the side chain sulfur atom of four invariant Cys residues. Rpp21 has most of its basic amino acid residues clustered around the N-terminal region of α₁, central loop, and the zinc ribbon domain at β₁. The electrostatic potential surface representation reveals a preponderance of positively charged on one face along the L-arm.

Proteins containing zinc-binding domains, such as Rpp21, have very diverse amino acid sequences and functions. For instance, the list includes RNA polymerase II initiation factor TFIIB, elongation factor TFIIS and eukaryotic translation initiation factor 2γ (Hahn et al, 2000; Olmsted et al, 1998; Roll-Mecak et al, 2004; Schmitt et al, 2002; Zhu et al, 1996). The only conserved sequence appears to be the two CXXC motifs. The sequence diversity among these protein, especially around the CXXC motifs, can be attributed to the fact that the only constraint is maintaining the planar nature of the zinc-binding fold. Such zinc-binding regions in proteins also eliminate the need for a hydrophobic core.
1.2.2.3.4 Rpp29

The structure of Rpp29 from *A. fulgidus* (*Afu*), *Mth* and *Pho* has been solved by using either NMR or X-ray diffraction methods (Fig. 1.5D; Boomershine et al, 2003; Numata et al, 2005; Sidote & Hoffman, 2003; Sidote et al, 2004). All these high-resolution studies reveal a conserved core structure of six antiparallel β-strands wrapped around a cluster of conserved hydrophobic residues. In both NMR studies (*Afu* and *Mth*), the C- and N-termini are unstructured (with only one helix determined in the C-terminal region of *Mth* Rpp29). However, in both crystal structures (*Afu* and *Pho*), there are well formed helices in the N- and C-termini. CD and NMR studies, the latter used to measure the rate of backbone amide protein exchange with the solvent, have shown that both termini in *Afu* Rpp29 are either unfolded in solution or in a rapid equilibrium between folded and unfolded states. The antiparallel β-sheet and β-barrel structures have been observed in the Sm family of RNA-associated proteins and the Hfq protein, a regulator of translation in bacteria (Schumacher et al, 2002).

Rpp29 is highly conserved among eukarya and archaea. There are several highly conserved hydrophobic residues located within the barrel structure. A conserved glycine residue contributes to a unique turn at the end of β2. Several solvent-exposed conserved residues could contribute to protein-protein and protein-RNA interactions. Although Rpp29 alone is not able to promote archaeal RPR catalytic activity and shows weak or no binding to archaeal RPR, yeast three-hybrid studies with subunits of both human and yeast RNase P have shown that Rpp29 interacts with its cognate RNA subunit (Jiang et al, 2001). NMR studies have shown that the presence of *Mth* RPR results in site-specific chemical shift perturbations in the 15N HSQC spectra of Rpp29 (Boomershine et al, 2003), specifically in
residues located at the $\beta_1$-$\beta_2$ loop and $\beta_6$. In addition, significant chemical shift perturbations were also observed in the flexible N- and C-terminal segments. Based on the chemical shift perturbations in the presence of its cognate RPR, and the electrostatic surface potential map of Rpp29, Numata et al. (2004) proposed two RNA-binding regions: (i) the concave surface, formed by both termini of Rpp29 rich in numerous basic residues, and (ii) the loop region connecting $\beta_2$ and $\beta_3$, based on similarity to Hfq (Boomershine et al., 2003; Numata et al., 2004) and the observation that Hfq uses these regions to interact with the A/U rich sequences in RNA by hydrogen bonding and stacking interactions (Schumacher et al., 2002). Although the Rpp29 termini are not rich in conserved residues, their importance is indicated by the observation that deleting 31 amino acid residues from the N-terminus of Pho Rpp29 resulted in a six-fold decrease in activity compared to the wildtype (Numata et al., 2004).

1.2.2.4 Protein-protein interactions in archaeal RNase P

Based on yeast two-hybrid analysis using subunits from human, yeast and archaeal RNase P, strong interactions between Pop5 and Rpp30 and between Rpp21 and Rpp29 were shown to be conserved in both archaeal and eukaryal Rpps (Jiang & Altman, 2001; Houser-Scott et al., 2002; Hall & Brown, 2004; Kifusa et al., 2005). Studies using NMR spectroscopy for the Pfu Rpp21+Rpp29 and Pop5+Rpp30 complexes revealed that there were significant chemical-shift perturbations in the HSQC spectrum of each member of each pair upon addition of its partner, indicating strong macromolecular interactions within the pairs (Boomershine, 2005; Wilson et al., 2006). Moreover, the human RNase P reconstitution studies have shown that human RPR shows weak activity when reconstituted
with Rpp21 and Rpp29 (Mann et al, 2003). Furthermore, yeast three-hybrid analyses indicated that human Rpp21, Rpp29 and Rpp30 individually interacted with the human RPR while yeast Rpp29 similarly interacts with its cognate RPR (Jiang et al, 2001). Despite information on possible interactions between various subunits of archaeal/eukaryal RNase P, hierarchical assembly maps are not available. In fact, this is one of the major reasons for not being able to reconstitute the archaeal/eukaryal RNase P holoenzymes.

1.3 Research Objectives

1.3.1 Bacterial RNase P

With a computer-aided model of the bacterial RPR and the tertiary structure of bacterial Rpp available, a major objective was to understand how this Rpp recognizes its cognate RPR to form a functional holoenzyme. Of course, an obvious extension would be to study the interaction of the substrate with the holoenzyme (i.e., RPR+Rpp). We believed that structural insights from reliable three-dimensional models of the RNase P holoenzyme would help appreciate how the sole bacterial Rpp, which is only one-tenth the size of bacterial RPR, plays various roles in RNase P catalysis, especially the increased affinity of the holoenzyme for ptRNA substrate. To achieve this goal, we decided to use a chemical modification method that specifically modifies Cys residues at select positions and thereby converts *E. coli* Rpp into a site-specific nuclease. By monitoring the cleavage sites of *E. coli* RPR that result from reconstitution with the modification introduced at different surface locations in *E. coli* Rpp, we sought to establish the proximity between specific nucleotides in *E. coli* RPR and certain amino acid residues in Rpp. To obtain such information between locations in *E. coli* Rpp (pre-assembled into a holoenzyme complex)
and the ptRNA substrate, a photochemical crosslinking method was employed. By specifically incorporating a photo-reactive group at defined positions in the central cleft of bacterial Rpp, we expected to deduce nucleotide positions in the ptRNA leader that are proximal to certain residues in E. coli Rpp. Taken together, we anticipated that these studies would furnish a significant number of distance constraints that would help in building three-dimensional models of the RNase P holoenzyme both in the absence and presence of its ptRNA substrate, a premise that has since been borne out (Chapter 2).

1.3.2 Archaeal RNase P

During the course of the last three years, the tertiary structures of all four archaeal Rpps have been solved by x-ray crystallography or NMR spectroscopy. A somewhat detailed secondary structure for the archaeal RPR is also available. However, establishing structure-function relationships in archaeal RNase P, which is a simpler and more tractable system than its eukaryal relative, has been thwarted by the lack of a robust reconstitution assay. In this regard, three major aims were undertaken in this dissertation: (i) optimize in vitro reconstitution of archaeal RNase P, under multiple turnover conditions, using Pyrococcus furiosus (Pfu) RNase P as our model system; (ii) determine if minimal functional cores do exist in Pfu RNase P and, if so, characterize their catalytic properties; and (iii) derive a hierarchical assembly map. Chapter 3 documents progress towards these objectives.
Figure 1.1 Role of RNase P in tRNA maturation. RNase P uses Mg$^{2+}$ to hydrolytically cleave the 5' leader sequence (red) of a ptRNA substrate to generate a mature tRNA. The cleavage by RNase P results in a 3' hydroxyl group in the leader sequence and 5' phosphate in the mature tRNA.
Figure 1.2 Classical representation of the 2D structure of the *E. coli* RNase P RNA (A) and of the *B. subtilis* RNase P RNA (B). Strands involved in the formation of the P4 and P6 pseudoknots are connected by square brackets. Extensions of stems P11, P10.1a, P15.1 and P19 are highlighted by continuous lines. Alternative representations for the 2-D structures of the RNase P RNA from *E. coli* and *B. subtilis*. These representations constitute an attempt at rendering the respective spatial arrangement of the helical stems and stacks as in the 3-D models. Thick lines link nucleotides adjacent in the sequence with arrows indicating 5' to 3' polarity. Nucleotides invariant among bacteria are shown in bold font. Nucleotides involved in tertiary interactions are enclosed in boxes and linked by thin, dotted lines. Known non-canonical pairings are indicated by open circles. For each 2-D diagram, the thick dotted line between P5 and P7 indicates the separation between the two main structural domains (Loria & Pan, 1996). Stems are color-coded according to the helical stacks proposed in the present 3D models (see Fig. 1.3). The figure and legend are reproduced and adapted, respectively, from Massire et al, 1998.
Figure 1.3 Comparison of the computer-aided three-dimensional models of bacterial RPRs and the tertiary structures recently established by X-ray crystallography. (A, B) Ribbon representations of the complete 3-D models of *E. coli* (Type A; panel A) and *B. subtilis* (Type B; panel B) RNase P RNA. The invariant nucleotides of the catalytic site are represented by large white spheres, while the conserved nucleotides from the T-loop recognition site are represented with smaller spheres. (C, D) Ribbon diagrams of the crystal structure of *T. maritima* (Type A; panel C) and *B. stearothermophilus* (Type B; panel D) RNase P RNA. The figures are reproduced from Kazantsev et al. (2005), Massire et al. (1998), and Torres-Larios et al. (2005).
Figure 1.4 Comparison of subunit composition of RNase P in all three domains of life. RNase P is composed of one essential RNA subunit and a varying number of protein subunits: one in bacteria, at least four in archaea, and up to ten in eukarya. The predicted secondary structures of the RNA subunits of RNase P from the three domains of life are depicted. The homology between protein subunits in archaeal and eukaryal RNase P is indicated by ~.
Figure 1.5 Tertiary structures of all four archaeal Rpps. (A) *Pfu* Pop5 (Wilson et al, 2006). (B) *Pho* Rpp30 (Takagi et al, 2004). (C) *Pho* Rpp21 (Kakuta et al, 2005) and (D) *Afu* Rpp29 (Sidote et al, 2004). *Pfu*-Pyrococcus furiosus; *Pho*-Pyrococcus horikoshii; *Afu*-Archaeoglobus fulgidus.
CHAPTER 2

MOLECULAR MODELING OF THE THREE-DIMENSIONAL STRUCTURE OF THE BACTERIAL RNASE P HOLOENZYME

2.1 Introduction

Ribonuclease P (RNase P) is an endoribonuclease that functions in precursor tRNA (ptRNA) processing in Bacteria, Archaea and Eukarya (Altman & Kirsebom, 1999; Frank & Pace, 1998; Gopalan et al, 2002; Hall & Brown, 2001). This study focuses on bacterial RNase P, a ribonucleoprotein (RNP) composed of one catalytic RNA subunit and one protein cofactor.

The *Escherichia coli* RNase P holoenzyme consists of M1 RNA (377 nucleotides, 123 kDa) and C5 protein (119 amino acid residues, 14 kDa). While M1 RNA cleaves ptRNAs to generate mature tRNAs in the absence of C5 protein in vitro (Guerrier-Takada et al, 1983), both subunits are essential for RNase P activity in vivo (Schedl & Primakoff, 1974). Several investigations have attempted to understand how M1 RNA, a metalloenzyme, mediates site-specific hydrolysis of a phosphodiester linkage in ptRNA (Christian et al, 2000; Harris et al, 1998; Warnecke et al, 1996). Specifically, mutagenesis, crosslinking and nucleotide analog interference mapping (NAIM) studies have identified nucleotides essential for ptRNA binding and catalysis.
Various studies have demonstrated that the protein cofactor, as part of the holoenzyme, facilitates substrate binding and enhances the rate of chemical cleavage (Crary et al, 1998; Kurz et al, 1998; Niranjanakumari et al, 1998; Peck-Miller & Altman, 1991; Tallsjo & Kirsebom, 1993). The tertiary structure of the protein subunit of RNase P from *Bacillus subtilis* and *Staphylococcus aureus* has recently been determined using X-ray crystallography and NMR spectroscopy, respectively (Fig. 2.1C; Spitzfaden et al, 2000; Stams et al, 1998). Also, a computer-aided three-dimensional model is available for the catalytic RNA moiety of bacterial RNase P (Chen et al, 1998; Massire et al, 1998). Nevertheless, the mechanism of action of the RNase P holoenzyme, especially the basis for protein-assisted RNA catalysis, has proved elusive due to the lack of high-resolution structural information on the RNA catalyst as well as the holoenzyme. In this study, we have used inferences from a hydroxyl radical-mediated footprinting approach together with findings on interactions in the RNase P-ptRNA complex reported elsewhere (Loria et al, 1998; Niranjanakumari et al, 1998; Svärd et al, 1996) to generate a list of the physical contacts between (i) the secondary structure elements in the two subunits of RNase P holoenzyme, and (ii) the holoenzyme and its ptRNA substrate. Using these data, we have derived a three-dimensional model of bacterial RNase P holoenzyme both in the absence and presence of its ptRNA substrate.
2.2 Materials and methods

2.2.1 Mutagenesis, purification, EPD-Fe modification and characterization of mutant derivatives of C5 protein

The four different single Cys-substituted mutant derivatives of C5 protein were generated using the QuickChange™ Site-Directed Mutagenesis Kit (Stratagene, San Diego, CA). The plasmid pBSC5Sn18 (which encodes the C5 C113S mutant derivative) was used as the template DNA for mutagenesis (Biswas et al, 2000). The following oligonucleotides and their antisense counterparts were employed for generating the mutants used in this study: T50C, 5' G GGG CAT CCC CGT ATC GGT CTC TGC GTC GCC AAG AAA AAC GTT CGA CGC G 3'; A52C, 5' CGT ATC GGT CTT ACA GTG TGC AAG AAA AAC GTT CGA CGC 3'; A59C, 5' GCC AAG AAA AAC GTT CGA CGC TGT CAT GAA CGC AAT CGG 3'; R62C, 5' CGA CGC GCC CAT GAA TGC AAT CGG ATT AAA CGT CTG ACG 3'; R70C, 5' CGG ATT AAA CGT CTG ACG TGT GAA TCC TTC CGT CTG CGC CAA CAT GAA CTC 3'. The boldface nucleotides indicate the codon that was altered to introduce the Cys residue. DNA oligonucleotides were synthesized by the Keck Biotechnology Resource Laboratory at Yale Medical School. Putative clones encoding the various mutant derivatives were sequenced to ensure that there were no additional unwanted alterations. Details of overexpression and purification of mutant derivatives in *E. coli* are provided elsewhere (Gopalan et al, 1997; Gopalan et al, 1999). After purification of the various mutant derivatives of C5 protein, we performed the modification with EPD-Fe as described by Biswas et al. (2000) and verified the molecular masses of the mutant derivatives, both before and after modification, using ESI mass spectrometric facilities at the
OSU Campus Chemical Instrumentation Center. Details on acquisition of the CD spectra as well as the RNase P assays are essentially as described in our earlier study (Biswas et al, 2000).

2.2.2 Footprinting experiments

The OH\(^{-}\)-mediated cleavages M1 RNA promoted by the EDTA-Fe modified derivatives of C5 protein were monitored in footprinting experiments exactly as described by Biswas et al. (2000).

2.2.3 Crosslinking assay and primer extension analysis

The crosslinking experiments were performed essentially as described by Niranjanakumari et al. (1998) with some modifications. First, for our studies with *E. coli* RNase P, we used an amber suppressor ptRNA\textsuperscript{Phe}, with a 10-nt leader sequence, as the substrate (Fig. 2.4A; Barrell & Sanger, 1969; McClain et al, 1987). Second, we reconstituted the RNase P holoenzyme using 200 nM of pre-folded M1 RNA and 500 nM of one of the single Cys-substituted mutant derivatives (e.g., V21C/C113S, T50C/C113S) either unmodified or modified with azidophenacyl bromide. Lastly, we used 15 nM of \(\alpha\)-(\(^{32}\)P)-GTP-labeled ptRNA\textsuperscript{Phe} in a final reaction volume of 10 µl.

The position in ptRNA\textsuperscript{Phe} which is crosslinked to Cys50 in C5 protein was mapped using a primer extension approach. The crosslinked complex (in lane marked T50C-AzP, Fig. 2.4B) was excised and first treated with 0.5 U proteinase K (Boehringer Mannheim) in 350 µl 20 mM Tris (pH 7.5), 100 mM NaCl, 0.01% (w/v) SDS and 1 mM EDTA. After an overnight incubation at 37\(^{\circ}\)C, the ptRNA in the
crosslinked complex was precipitated and subjected to reverse transcription with 15 fmol of a 5' end-labeled oligonucleotide primer (5' GCCCGGATCGGAATCG 3') that is complementary to the 3' end of ptRNA\textsuperscript{Phe}. Primer extension was performed according to the manufacturer’s instructions for the Thermoscript\textsuperscript{TM} RT-PCR system (Invitrogen Gibco Life Technologies). The extension products and the sequencing ladder for ptRNA\textsuperscript{Phe} were then separated using 8% (w/v) polyacrylamide/7M urea gel electrophoresis.

2.2.4 Molecular modeling

The 3D models of the RNase P RNAs from \textit{E. coli} and \textit{B. subtilis} were interactively built as described (Westhof & Michel, 1994; Massire et al, 1998) either naked or bound to the ptRNA and to the protein using the software MANIP (Massire & Westhof, 1998). The docking of the protein was performed with the goal of ensuring that the nucleotides cleaved by hydroxyl radicals were in the vicinity of corresponding modified Cys residues. Such an exercise cannot be expected to yield an accurate RNA-protein complex model in terms of the intra-molecular interactions. The RNA regions of the models were subjected to restrained geometrical least-square refinement using the program NUCLIN/NUCLSQ (Westhof, 1993) to ensure geometry and stereochemistry with correct distances between interacting atoms and to avoid steric clashes. The color illustrations were generated with the program DRAWNA (Massire et al, 1994). The coordinates for the three-dimensional models are available upon request.
2.3 Results

2.3.1 Hydroxyl radical-mediated footprinting approach

Various single Cys-substituted mutant derivatives of C5 protein were modified using an SH-specific iron complex of EDTA-2-aminoethyl 2-pyridyl disulfide (EPD-Fe; Fig. 2.1D). This chemical reaction converts the mutant derivatives of C5 protein into chemical nucleases which, when reconstituted as part of the RNase P holoenzyme, will cleave their cognate RNA ligand in the presence of ascorbate and hydrogen peroxide (Ermacora et al., 1992; Ermacora et al., 1994; Hall & Fox, 1999). Upon reduction of the iron (covalently tethered to a Cys residue) with ascorbate, reactive oxygen species are generated and oxidative degradation of the M1 RNA backbone proximal to the modified residue occurs. The chemical cleavages are usually localized and restricted to < 10 Å of the metal ion because the reactive OH· species has a very short lifetime in aqueous solution (Hall & Fox, 1999; Joseph et al., 1997). Since the Fe atom is 14 Å from the Cα position of the Cys residue to which the EDTA-Fe moiety is conjugated, this OH·-mediated footprinting must be viewed as a long-range structural probing method. Success of EDTA-Fe-based, site-specific footprinting strategies depends on the modified Cys residue being proximal to the RNA-protein interface without causing any deleterious effects on the assembly of the RNP complex (Hall & Fox, 1999). We already have had success in using this methodology for mapping contact sites in the *E. coli* RNase P holoenzyme (Biswas et al., 2000). Due to insufficient information on the distance constraints both with respect to the RNP interactions in the holoenzyme as well as the holoenzyme-ptRNA substrate contacts, we have now extended this earlier study.
2.3.2 EDTA-Fe modification and characterization of single Cys-substituted mutant derivatives of C5 protein

Based on a sequence alignment of the protein subunit of RNase P from different bacteria, we selected four positions that are either at or proximal to the highly conserved residues of the RNR motif in helix α2 or in the loop preceding this motif (Figs. 2.1B and C). The residues Ala52, Ala59, Arg62 and Arg70 in C5 protein (Fig. 2.1C) were individually mutated to Cys and modified with EPD-Fe to covalently tether EDTA-Fe to the thiol group (Fig. 2.1D). In these four mutants, Cys113, which is present in the wild type C5 protein, was mutated to Ser. To verify (i) the molecular masses of the four different single Cys-substituted mutant derivatives that we purified, and (ii) if site-specific attachment of EDTA-Fe to the unique thiols in these mutants was accomplished, electrospray ionization mass spectrometry (ESI-MS) was performed on the mutants before and after modification. The molecular masses obtained from ESI-MS confirmed the identity of the mutants as well as the successful and complete modification with EPD-Fe (data not shown). Subsequently, we used circular dichroism (CD) spectroscopy to investigate if modification of the four mutant derivatives resulted in gross structural alterations. The CD spectra of the mutants (± EDTA-Fe) were similar to that observed for the wild type protein (data not shown).

To examine whether the single Cys-substituted mutant derivatives, with and without the EDTA-Fe tether, can form a functional RNase P holoenzyme, they were reconstituted with wild type M1 RNA and the initial velocity for cleavage of ptRNA^Tyr substrate measured (Table 2.1). While the mutation to Cys per se is not detrimental for RNase P activity, attachment of EDTA-Fe to the thiol does cause a decrease in
activity depending on the site of modification. For example, C5 A52C/C113S before and after modification displayed 81% and 27%, respectively, of the activity observed with the wild type RNase P holoenzyme. However, the modification in the four different mutants never resulted in activity that was less than 25% of that observed with the wild type RNase P holoenzyme (Table 2.1).

Based on the results described above, we conclude that the four single Cys-substituted mutant derivatives of C5 protein are folded and functional, in the absence and presence of a EDTA-Fe tether. Nevertheless, since the positions chosen for Cys mutagenesis and subsequent modification are at or near locations of highly conserved residues, it is not surprising that the modification causes a decrease in activity (see Discussion).

2.3.3 Footprinting studies

All modified and unmodified mutant derivatives of C5 protein were reconstituted with either 5’ or 3’ end-labeled M1 RNA, and the holoenzymes immediately incubated with ascorbate and hydrogen peroxide. To map the exact positions of OH·-mediated cleavages in M1 RNA, the footprinting reaction contents were separated using either 8% or 10% polyacrylamide/7M urea gel electrophoresis and the cleavage sites identified by autoradiography as well as phosphorimager analysis. Using 5’ end-labeled M1 RNA, OH·-mediated cleavages were observed with C5 Cys52- and Cys59-EDTA-Fe but not with C5 Cys62- or Cys70-EDTA-Fe. While both C5 Cys52- and Cys59-EDTA –Fe promoted OH·-mediated cleavages at positions 67 and 68 in M1 RNA (Fig. 2.2), weak cleavages were also observed from positions 69 through 72 with C5 Cys52-
EDTA-Fe. In contrast, using 3’ end-labeled M1 RNA, OH-mediated cleavages were observed with C5 Cys62- and Cys70-EDTA-Fe but not with C5 Cys52- or Cys59-EDTA-Fe. Cys62-EDTA-Fe promoted OH-mediated cleavages at positions 331, 332, 349, 352, and, albeit weakly, 350; Cys70-EDTA-Fe caused scission with moderate intensity at positions 331 and 332 (Fig. 2.2). The footprints of seven different single Cys-substituted mutant derivatives (three from an earlier study and four from this investigation) are restricted to domain 2 of M1 RNA, where the active site and most of the conserved nucleotides are located (Fig. 2.3).

The 14-Å arm of the EDTA-Fe moiety makes it difficult to precisely pinpoint the nucleotides in M1 RNA that are interacting with C5 protein. However, the latitude afforded by the diffusibility of the hydroxyl radicals ensures the identification of a range of possible sites of interaction between the two subunits in the RNase P holoenzyme. We recognize that the validity and accuracy of the footprinting data should be ascertained with rigor. In this regard, various observations from our studies are worthy of mention. First, the footprints were highly reproducible and verified in three or more independent trials. Second, since the signals (i.e., the OH-mediated cleavages) are frequently weak and sometimes obscured by background noise (i.e., cleavages caused by the conditions used for denaturation and annealing of the RNA moiety), it is reassuring to note that not all modified derivatives resulted in a footprint; for example, Cys106- and Cys113-EDTA-Fe failed to promote OH-mediated cleavages of M1 RNA, likely due to these residues being distal from the RNA-protein interface in the RNase P holoenzyme (Biswa et al, 2000; data not shown). Third, the OH-mediated cleavages were never identical for any two EDTA-Fe-tethered derivatives of C5 protein although overlap in the footprint
was certainly observed (Fig. 2.2); the latter finding could be accounted for by the spatial proximity of the sites of modification (Fig. 2.1). Fourth, the footprints of C5 Cys52- and Cys59-EDTA-Fe were at identical locations in the RNA subunits of both E. coli and Neisseria gonorrhoeae RNase P (data not shown). Last, the footprint of C5 Cys16-EDTA-Fe in the P3 helix of M1 RNA is corroborated by the results from a crosslinking experiment which indicated that this helix is proximal to C5 protein in the RNase P holoenzyme (Guerrier-Takada, C. & Altman, S., personal communication).

2.3.4 Non-covalent interactions in the ES complex

RNase P-mediated cleavage occurs between the –1 and +1 positions of the ptRNA substrate. Recent experiments with single Cys-substituted mutant derivatives of the protein subunit of B. subtilis RNase P have revealed that the leader sequence of the ptRNA<sup>Asp</sup> interacts with the cleft in the protein subunit. For instance, a strong crosslink was observed between the -4 and -8 nucleotides in the leader sequence of the ptRNA substrate and residues 49 and 27 of the protein subunit, respectively (Niranjanakumari et al, 1998). Consistent with the expectation that the protein cofactor might enhance substrate binding by promoting direct contacts with the leader sequence, determination of the binding affinities of the B. subtilis RNase P holoenzyme for ptRNA<sup>Asp</sup> with different leader lengths revealed a precipitous decrease in affinity when the leader was shortened to four nucleotides or less (Crary et al, 1998).

To facilitate molecular modeling of the E. coli RNase P holoenzyme-substrate ternary complex, we extended the study by Niranjanakumari et al. (1998) and established that the protein subunit of E. coli RNase P binds the ptRNA substrate in a
manner identical to \textit{B. subtilis} RNase P and that this interaction is independent of the identity of the ptRNA (Fig. 2.4; data not shown). Of relevance to this study is the finding that the T50C mutant derivative of C5 protein when modified with the photoactivatable azidophenacyl bromide (APBr) and reconstituted with M1 RNA yielded a strong UV-induced crosslink at the –4 and –5 nucleotides in the leader sequence of ptRNA$^{\text{Phe}}$ (Fig. 2.4). This information was vital in building a model of the \textit{E. coli} RNase P-ptRNA complex.

\textbf{2.3.5 Computer-aided molecular modeling}

The footprinting data were used to guide us in positioning the protein subunit on the catalytic RNA moiety. Since the tertiary structure of the \textit{E. coli} RNase P protein subunit has not been determined, the high resolution structure of the \textit{B. subtilis} RNase P protein cofactor was used for building the \textit{E. coli} RNase P holoenzyme model. Since the two proteins share nearly 30\% identity, their tertiary structures are expected to be similar. In fact, this expectation is borne out by modeling/threading exercises (Sharkady \& Nolan, 2001; Biswas \& Gopalan, unpublished data). Perhaps more importantly, the observation that heterologous reconstitution of RNase P activity can be accomplished using non-cognate bacterial RNase P RNA and protein subunits (e.g., \textit{B. subtilis} RNA + \textit{E. coli} protein; Guerrier-Takada et al, 1983) lends merit to our docking approach in which the tertiary structure of the \textit{B. subtilis} protein cofactor is placed on the three-dimensional model of either the \textit{E. coli} or \textit{B. subtilis} RNase P RNA subunit using distance constraints between specified locations in the two subunits of \textit{E. coli} RNase P.
Sequence comparisons were used to establish the amino acid positions in the *B. subtilis* RNase P protein subunit which correspond to the locations in C5 protein that were covalently attached to EDTA-Fe (Figs. 2.1A and C). Along with the three mutant derivatives of C5 protein reported in Biswas et al. (2000), we have seven positions where the EDTA-Fe moiety was introduced. These positions encompass three distinct regions of the protein (Fig. 2.1C) and describe a plane that can be fitted onto the RNA model. The footprinting data obtained using EDTA-Fe tethered to these positions provided sufficient distance constraints to dock the protein cofactor on domain 2 of the RNA subunit and help build nearly identical three-dimensional models of the *E. coli* and *B. subtilis* RNase P holoenzymes (Fig. 2.5; data not shown).

While docking of the protein is facilitated by the distance constraints yielded by the footprinting method, it is important to ensure that any holoenzyme model generated using these data (Fig. 2.5) should also permit the construction of a tertiary structure of the holoenzyme-substrate (ES) complex in which well-established physical contacts between the enzyme and substrate are easily rationalized. Three different interactions believed to occur in the RNase P holoenzyme complex were taken into consideration while building the three-dimensional model of the RNase P-ptRNA complex (Fig. 2.6): (i) the base pairing between the 3’-terminus of the ptRNA (C75 and C76) and loop L15 (G292 and G293) of M1 RNA (Kirsebom & Svärd, 1994; Oh & Pace, 1994; Svärd et al, 1996), (ii) the contacts between functional groups in the T stem-loop of the ptRNA substrate with the paired regions P9 and P11 of the catalytic RNA moiety (Loria et al, 1998), and (iii) the binding of the ptRNA leader (nucleotides -4 to –8) in the cleft of the protein cofactor (Niranjanakumari et al, 1998; Fig. 2.4).
Molecular modeling was initiated using the previously reported three-dimensional models of RNase P RNAs from *E. coli* and *B. subtilis* (Massire et al, 1998). We soon realized that the protein could be docked on these earlier models only if it were positioned to interact with the RNase P RNA on the side opposite to the one that binds the ptRNA. However, this would have been inconsistent with the experimental data that indicate direct contacts between the ptRNA leader sequence and the protein subunit of both *E. coli* and *B. subtilis* RNase P. This type of discrepancy was at least partly expected since the previous models were built primarily on the basis of results from phylogenetic covariation analysis and data obtained from experiments that involved only the RNA subunit. Nevertheless, without drastically altering the overall architecture of the previous model of the RNA subunit, especially the positioning of various highly conserved nucleotides proximal to the cleavage site, we were able to build a model of the RNase P holoenzyme by employing the following approach.

The P15 helix in the earlier model (Massire et al, 1998) was too close to the substrate and was virtually taking the place the protein would be expected to occupy based on the footprinting and crosslinking data. Therefore, a cavity was created to accommodate the protein in the vicinity of the ptRNA leader sequence by pushing P15 away from the leader sequence (Fig. 2.6). This was achieved by first reorienting P5 using the phosphate group of C242 as a pivot without disrupting its connection to P4. Such a modification made possible the *en bloc* movement of P6, P15, P16, and P17 helices as a rigid body and the creation of a cavity necessary to accommodate the protein. Placement of the protein subunit in this cavity immediately reconciled both the footprinting and crosslinking data (Fig. 2.6A). Note that the antiparallel helix formed between loop
L15 (G292 and G293) and the 3’ terminus of the ptRNA (C76 and C75) is maintained in this model (Fig. 2.6B).

2.4 Discussion

In this study, we have employed a site-specific footprinting approach to map potential contact sites between M1 RNA and C5 protein in the *E. coli* RNase P RNP complex. Furthermore, this information has been utilized to generate a three-dimensional model of the bacterial RNase P holoenzyme in the presence and absence of its ptRNA substrate.

2.4.1 Directed hydroxyl radical-mediated footprinting strategy

Four different single Cys-substituted mutant derivatives of C5 protein were constructed to facilitate the strategic placement of an affinity cleavage reagent at positions which are likely to be at the RNA-protein interface in the *E. coli* RNase P holoenzyme. The attachment of EDTA-Fe to these mutant derivatives resulted in activities ranging from 27 to 42% of that observed with wild type RNase P (Table 2.1). The loss of activity upon modification is to be expected since the thiols were engineered at locations that are either at or close to conserved residues. As indicated in our earlier study (Biswas et al, 2000), these results highlight a limitation with this footprinting approach. Specifically, to gain insights into RNA-protein interactions in an RNP complex, the RNA-binding protein is converted into a site-specific chemical nuclease by introducing a cleavage reagent at the putative RNA-protein interface (Hall & Fox, 1999). However, such a chemical modification at a site critical for RNP assembly
could potentially result in loss of biological activity. Nevertheless, the strategy would be expected to yield valuable structural insights if there is reasonable residual activity after modification. Based on this premise and the recent finding that OH-mediated footprinting data on the contact sites between S5 and 16 S rRNA have been borne out by the high-resolution structural studies of the 30 S ribosomal subunit (Heilek & Noller, 1996; Brodersen et al, 2002), we proceeded with the footprinting study on *E. coli* RNase P.

### 2.4.2 Location of the protein-binding site in the RNA subunit of bacterial RNase P

The footprinting data from this and our previous study (Biswas et al, 2000) can be summarized as follows: helix α1, the unique left-handed crossover loop, and the RNR motif in helix α2 of C5 protein are proximal to the P3, P4 and J18/2 regions of M1 RNA, respectively; some residues in helix α2 are also close to J2/4. Interestingly, results from crosslinking and NAIM methods have already established the vital role of several nucleotides that are part of P4, J18/2 and J2/4 (in M1 RNA) during RNase P RNA-mediated catalysis (Christian et al, 2000; Kaye et al, 2002; Kufel & Kirsebom, 1996; Siew et al, 1999).

A phylogenetic covariation analysis of 137 bacterial RNase P RNA sequences resulted in the generation of a secondary structure model which is composed of distinct substrate specificity (domain 1) and catalytic (domain 2) regions (Fig. 2.3; Massire et al, 1998). The OH-mediated footprint from all seven single Cys-substituted mutant derivatives of C5 protein is restricted to domain 2 of M1 RNA (Fig. 2.3). Using the
footprinting data, we have docked the protein cofactor on M1 RNA and constructed a model of the RNase P holoenzyme (Fig. 2.5).

One of the most important issues in theoretical modeling relates to the methods employed for validation. We considered a quantitative approach to verify the accuracy of the docking model. Since the linker arm in EDTA-Fe positions the Fe atom 14 Å from the Cα position in the modified Cys residue (of C5 protein) and the short half-life of the OH- does not permit its diffusion 10 Å past the Fe atom, the distances between the Cα position and the phosphorus centers where OH- mediated cleavages have occurred in M1 RNA must be approximately 14 ± 10 Å. Iterative manual building/refinement cycles of the model generated a set of Cα-PO4 distances in the E. coli RNase P holoenzyme complex that were consistent with this expectation (Table 2.2; see below). Considering that far greater distances have been reported for RNA-protein interactions probed by this method in 16S rRNA (Whirl-Carillo et al, 2002), we were encouraged that three-fourths of the Cα-PO4 distances fulfilled the 14 ± 10 Å criterion (Table 2.2). But what about the remainder that failed this test? In our modeling approach, no allowance has been made for conformational changes that might occur in either M1 RNA or C5 protein during RNase P assembly. The largest discrepancies in Cα-PO4 distances were noted with the Cys residues located in the loop connecting strand β3 and helix α2 (Fig. 2.1; Table 2.2). The location of this loop in the unique left-handed crossover motif, which plays an important biological role in other proteins where it is present (Stams et al, 1998), supports the notion that this loop in C5 protein might undergo structural rearrangement upon binding to M1 RNA. Based on our footprinting data, this loop was placed proximal
to the P4 helix and the adjoining single-stranded regions (J3/4), which contain several universally conserved nucleotides (Massire et al., 1998). NAIM and site-specific phosphorothioate substitution followed by thiophilic metal ion rescue experiments have led Harris and coworkers to conclude that multiple metal ions, essential for catalysis, interact with specific base functional groups in P4 as well as J3/4 (Christian et al., 2000; Christian et al., 2002; Kaye et al., 2002). Since C5 protein can diminish the requirement for Mg$^{2+}$ during M1 RNA-mediated catalysis (Guerrier-Takada et al., 1983), it is conceivable that specific amino acid residues (at or proximal to the loop connecting strand β3 and helix α2) in C5 protein might directly or indirectly facilitate tighter metal ion binding by the P4 helix. High-resolution structural studies are needed to elucidate the dynamics and metal-binding properties of this loop, which might be critical for RNase P assembly and function.

### 2.4.3 Is the holoenzyme model consistent with previous data?

We discuss below our model of the RNase P holoenzyme in the context of findings from other studies that have attempted to elucidate RNA-protein interactions in bacterial RNase P.

The binding site of C5 protein in domain 2 (P3, P4, J18/2, and J2/4) of M1 RNA is perhaps not unexpected since domain 2, which contains most of the conserved residues (white spheres in Fig. 2.5C), is sufficient for catalysis under certain in vitro conditions (Guerrier-Takada & Altman, 1992; Pan, 1995). For example, the observation that the protein cofactor can restore activity in vitro to a deletion variant of the RNA moiety encompassing domain 2 alone (Green et al., 1996; Guerrier-Takada & Altman, 1992) is
consistent with our observation that the binding site for the protein subunit is restricted to this catalytic domain (Fig. 2.3).

Recently, Sharkady and Nolan (2001) identified RNA-protein contact sites in *E. coli* RNase P by employing a photochemical crosslinking approach using azidophenacyl-conjugated single Cys-substituted mutant derivatives of C5 protein. The sites of crosslinking (P5-J5/P15, P11, J11/12, P12, J14/11) and the three-dimensional structure of the *E. coli* RNase P holoenzyme based on these data are quite different from our docking model (Fig. 2.4; see discussion below for possible reasons underlying this disagreement).

Both chemical and enzymatic probes have also been employed to deduce RNA-protein interactions in bacterial RNase P. A recent study utilized NAIM to determine the positions in the RNA subunit of *B. subtilis* RNase P at which introduction of AMPαS and IMPαS modifications interfered with binding to the cognate protein subunit (Rox et al, 2002). Indeed, a few modifications in domains 1 and 2 of the RNA subunit weakened binding to the protein cofactor. It is notable that positions identified by Rox et al. (2002) constitute a subset of those identified in another investigation on *B. subtilis* RNase P wherein the putative protein binding sites on the RNA were deduced based on protein-induced protections of the RNA backbone to OH-‐mediated cleavages (Loria et al, 1998). There is excellent agreement between the elements in domain 2 identified as essential for RNA-protein interactions by the NAIM study and the regions that we have determined here as the putative protein-binding site in the RNA (Fig. 2.3). However, it is important to note that our study failed to identify a protein footprint in domain 1, in contrast to findings from two earlier studies on *E. coli* (Talbot & Altman, 1994; Vioque et al, 1988) and *B. subtilis* RNase P (Loria et al, 1998; Rox et al, 2002). Structural
alterations in regions of the RNA distal to the protein binding site are not uncommon and complicate inferences on the RNA-binding site using various chemical and enzymatic probes. For instance, in the NAIM or protection from RNase T1 studies (Rox et al., 2002; Vioque et al., 1988), if modifications in domain 1 alter the structure of domain 2 and thereby diminish protein binding, an interference detrimental for protein binding would be identified in domain 1 even if domain 1 is not directly contacted by the protein cofactor. Similarly, in the OH·-mediated footprinting study, the binding of the protein subunit to domain 2 could result in changes in the susceptibility of the RNA backbone in domain 1 to OH·-mediated cleavages and lead to the conclusion that the protein contacts both domains. Clearly, the data from these experiments must be interpreted with caution.

The above mentioned results lend ambiguity to the placement of the protein cofactor on M1 RNA and warrant further discussion. Since the small protein subunit can only bridge 40 Å across its RNA ligand, Rox et al. (2002) concluded that simultaneous contacts with the two domains of the RNA subunit by a single protein molecule was rather unlikely (at least based on the current models of the RNA). Rox et al. (2002) rationalized that domain 1 might constitute a protein-binding site independent of domain 2 consistent with a model in which RNase P is an RNP complex comprising two copies each of the RNA and protein subunit. Indeed, a symmetrical tetrameric structure for B. subtilis RNase P was recently proposed on the basis of small angle X-ray scattering (SAXS) and affinity retention experiments (Fang et al., 2001). However, there are no data regarding the subunit stoichiometry of bacterial RNase P in vivo.

The differences between our footprinting data and the results of Sharkady and Nolan (2001) and Rox et al. (2002) might stem from differences in the oligomeric
state of the enzyme, which in turn is likely influenced by the concentration of enzyme. Our footprinting experiments were performed using holoenzyme concentrations lower than those used for the crosslinking, NAIM and SAXS studies. This fact becomes especially relevant considering that (i) a Scatchard analysis of the interaction between C5 protein and M1 RNA revealed non-ideal binding behavior presumably due to the presence of high-affinity and low-affinity binding modes (Talbot & Altman, 1994), and (ii) C5 protein at high concentrations can non-specifically bind to RNAs other than M1 RNA (Vioque et al, 1988). A rationale for concentration-dependent formation of the different oligomeric states of bacterial RNase P might yet provide a basis for explaining the disagreement in the final three-dimensional models derived from results of various technical approaches. Of course, the validity of any model can also be assessed by examining the extent to which it is collectively consistent with a large body of experimental data.

2.4.4 Three-dimensional model of the ES complex

The tertiary structure model of the RNase P holoenzyme-ptRNA complex (Fig. 2.6) presented here is compatible with various interactions that have already been demonstrated elsewhere. First, several studies have now confirmed that the two C’s in the 3’-CCA motif of the ptRNA substrate interact with the highly conserved G292 and G293 in M1 RNA in both the RNA- and holoenzyme-catalyzed reactions (Heide et al, 1999; Kirsebom & Svärd, 1994; Oh & Pace, 1994). Although this interaction occurs in RNase P from several bacteria (Svärd et al, 1996), these canonical Watson-Crick base pairs are not likely to be universal since the nucleotides at 292 and 293 in the RNA subunit
of RNase P are not absolutely conserved. Second, data from circular permutation analysis and photochemical crosslinking indicate that specific nucleotides in P9 and P11 (domain 1) of M1 RNA contact the T stem-loop of the ptRNA substrate. Although most of these studies were performed with the catalytic RNA in the absence of the protein cofactor, one study using the \textit{B. subtilis} RNase P holoenzyme supports the idea of a specific hydrogen-bonding interaction between the N1 of A230 and the 2\’ hydroxyl group at position 60 in the T stem-loop of the ptRNA (Loria et al, 1998). Lastly, the binding of the ptRNA leader (nucleotides -4 to –8) in the cleft of the protein cofactor has been demonstrated by photo-crosslinking studies (Niranjanakumari et al, 1998; this study). These interactions collectively provided constraints to position the 5\’, T stem-loop and 3\’ ends of the ptRNA (Fig. 2.6).

Once the model of the RNase P holoenzyme in the presence of its ptRNA substrate was constructed so as to fulfill the above mentioned constraints (Fig. 2.6), some other important features became evident. First, the two highly conserved residues A248 and A249 (in J5/15 of M1 RNA) lie in the shallow groove of the acceptor stem such that their Watson-Crick base-pairing edges contact the shallow groove edges of nucleotides –1 and –2 of the ptRNA substrate. A direct role for A248 and A249 in substrate recognition, especially close to the cleavage site, was already predicted based on (i) efficient short-range crosslinks between A248 and the +1 position of the ptRNA, and (ii) significant interference effects in binding and catalysis upon modification of A248 and A249 to adenosine base analogues (Christian et al, 1998; Christian & Harris, 1999; Siew et al, 1999). Second, the nucleotide –1 of the ptRNA is close to the universally conserved A352 while other residues from the P4 helix also appear to be ideally positioned for
recognition of the acceptor stem of the ptRNA substrate. The recent identification of a polynuclear metal ion-binding site at the P1-P4 multihelix junction provides further support for P4 and its proximal elements as being part of the catalytic core (Christian et al, 2000; Christian et al, 2002). Lastly, the highly conserved RNR motif of helix α2 (Fig. 2.1B) is not close to the scissile phosphodiester linkage in the ptRNA substrate consistent with results from the crosslinking study with *B. subtilis* RNase P which established contacts between the ptRNA leader (nucleotides -4 to -8) and residues in the cleft but not with residues in the RNR motif of helix α2 in the protein subunit (Niranjanakumari et al, 1998). However, an indirect participation of the highly conserved RNR motif in RNase P catalysis cannot be ruled out. By interacting with conserved nucleotides in J18/2, the RNR motif could facilitate substrate recognition and/or assist in metal ion binding at the crossover loop.

### 2.4.5 Mechanism of action of bacterial RNase P and evolutionary considerations

The three-dimensional model of the ES complex provides a structural basis for the role of the protein cofactor, a central theme in several recent studies (Fig. 2.6). First, the ability of C5 protein to bind the distal part of the leader sequence (nucleotides -4 to -8) while simultaneously interacting with conserved nucleotides in domain 2 of M1 RNA, which play a vital role in holoenzyme-mediated catalysis, might explain how the protein cofactor can contribute towards enhancing both substrate binding and the rate of chemical cleavage (Crary et al, 1998; Kurz et al, 1998). Moreover, the model postulates that C5 protein employs different amino acid residues to accomplish concomitant (but distinct) interactions with the catalytic RNA moiety and the ptRNA substrate;
specifically, the conserved basic residues in the RNR motif of helix α2 bind conserved regions in M1 RNA while residues in the β-sheet and helix α1 help bind the leader sequence in the substrate (Figs. 2.1 and 2.6). Site-directed mutagenesis together with a genetic complementation assay has already furnished some insights with regard to the possible dual RNA-binding characteristics in C5 protein (Jovanovic et al, 2002). Second, it appears that there is exquisite cooperativity between the two RNase P subunits in ptRNA substrate binding. For instance, M1 RNA utilizes P9, P11 and L15/16 to make specific contacts with the T stem-loop as well as the 3′ end of the acceptor stem in the ptRNA, while the protein contacts the distal end of the ptRNA leader sequence. Entropic constraints imposed by multiple and concomitant interactions of the ptRNA with both subunits of bacterial RNase P likely contribute significantly to substrate recognition/metal-ion binding and ensure a catalytic rate which approximates a diffusion controlled reaction (Kurz et al, 1998).

If the mechanism of action of the holoenzyme (with respect to ptRNA processing) is identical to that observed with the RNA subunit alone (Warnecke et al, 1997; Loria et al, 1998), why was addition of the protein cofactor required? An already energetically unfavorable interaction between M1 RNA and ptRNA could have been exacerbated gradually by alterations in the cellular milieu (particularly the ionic conditions). Based on a large body of data, it has already been postulated that C5 protein was probably recruited to aid M1 RNA by enhancing substrate binding and product release, and by decreasing the requirement of metal ions essential for catalysis (Crary et al, 1998; Gopalan et al, 1994; Kurz et al, 1998; Reich et al, 1988; Tallsjo & Kirsebom, 1993). Since proteins can offer coordinating ligands for tight binding of metals, it is
possible that the affinity for the metal ions improved considerably in the holoenzyme, thereby ensuring that a metalloenzyme like M1 RNA could function at lower concentrations of metal ions (like Mg\(^{2+}\)). Such metal ion-binding sites on the protein cofactor remain to be identified. The expectation that the protein merely acts as an electrostatic shield that promotes the interaction of two RNAs has been replaced with a molecular mechanism elaborated by Fierke and coworkers (Kurz & Fierke, 2000). By establishing direct, non-covalent interactions with the ptRNA leader sequence, the protein engenders a favorable ribozyme-ptRNA ground state interaction. One would predict on the basis of this scenario that recruitment of the protein cofactor either replaced or eliminated a subset of the entire suite of interactions between the RNA subunit and the ptRNA substrate. For instance, in the holoenzyme model (Fig. 2.5), distal parts of the ptRNA leader sequence are docked in the cleft of the protein cofactor and cannot be proximal to J5/15 or J18/2 as established by studies on the M1 RNA-ptRNA complex (Christian & Harris, 1999). Experiments designed to test postulates of this nature might yield results central to understanding of the mechanism of protein-assisted RNA catalysis in bacterial RNase P and also furnish insights into the likely steps during the progression from an RNA to an RNP enzyme.
Table 2.1 Relative RNase P activities of various mutant derivatives of C5 protein (before and after EPD-Fe modification) and a summary of their footprinting results.

<table>
<thead>
<tr>
<th>Mutant derivative of C5 protein</th>
<th>Relative initial velocity (%)*</th>
<th>Positions in M1 RNA cleaved by directed hydroxyl radicals</th>
</tr>
</thead>
<tbody>
<tr>
<td>C5 A52C/C113S</td>
<td>81  27</td>
<td>67, 68, 69, 70, 71, 72</td>
</tr>
<tr>
<td>C5 A59C/C113S</td>
<td>65  28</td>
<td>67,68</td>
</tr>
<tr>
<td>C5 R62C/C113S</td>
<td>60  30</td>
<td>331, 332, 348, 349, 350, 352</td>
</tr>
<tr>
<td>C5 R70C/C113S</td>
<td>72  42</td>
<td>331,332</td>
</tr>
</tbody>
</table>

* Relative initial velocity is normalized based on the activity of the wild type C5 protein which is considered as 100%. The turnover number of the wild type RNase P holoenzyme was 16 min$^{-1}$. The initial velocity data were calculated from three independent trials with the standard deviation not exceeding 17%.
<table>
<thead>
<tr>
<th>AA</th>
<th>Nt</th>
<th>Distance (Å)</th>
<th>AA</th>
<th>Nt</th>
<th>Distance (Å)</th>
<th>AA</th>
<th>Nt</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ser16</td>
<td>C29</td>
<td>22.6</td>
<td>Ala52</td>
<td>A67</td>
<td>31.1</td>
<td>Arg62</td>
<td>U331</td>
<td>10.2</td>
</tr>
<tr>
<td>G30</td>
<td>18.8</td>
<td></td>
<td>G68</td>
<td>27.0</td>
<td></td>
<td>G332</td>
<td>12.7</td>
<td></td>
</tr>
<tr>
<td>U31</td>
<td>14.4</td>
<td></td>
<td>U69</td>
<td>24.0</td>
<td></td>
<td>C348</td>
<td>22.7</td>
<td></td>
</tr>
<tr>
<td>C32</td>
<td>10.0</td>
<td></td>
<td>C70</td>
<td>24.2</td>
<td></td>
<td>A349</td>
<td>16.2</td>
<td></td>
</tr>
<tr>
<td>G33</td>
<td>11.0</td>
<td></td>
<td>C71</td>
<td>25.2</td>
<td></td>
<td>G350</td>
<td>11.7</td>
<td></td>
</tr>
<tr>
<td>A49</td>
<td>15.2</td>
<td></td>
<td>Lys54</td>
<td>A66</td>
<td>34.4</td>
<td>A351</td>
<td>5.2</td>
<td></td>
</tr>
<tr>
<td>C50</td>
<td>11.9</td>
<td></td>
<td>A67</td>
<td>29.4</td>
<td></td>
<td>A352</td>
<td>5.6</td>
<td></td>
</tr>
<tr>
<td>G51</td>
<td>12.1</td>
<td></td>
<td>G72</td>
<td>19.5</td>
<td></td>
<td>Lys66</td>
<td>G332</td>
<td>9.6</td>
</tr>
<tr>
<td>G52</td>
<td>15.7</td>
<td></td>
<td>G73</td>
<td>19.4</td>
<td></td>
<td>G333</td>
<td>11.5</td>
<td></td>
</tr>
<tr>
<td>Ala59</td>
<td>A74</td>
<td>21.2</td>
<td>G74</td>
<td>21.2</td>
<td></td>
<td>C348</td>
<td>24.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A67</td>
<td>20.3</td>
<td></td>
<td>68</td>
<td>16.9</td>
<td>A351</td>
<td>7.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arg70</td>
<td>U331</td>
<td>13.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G332</td>
<td>9.8</td>
</tr>
</tbody>
</table>

Table 2.2 Cα-PO₄ distances in the model of *E. coli* RNase P ternary complex

* The length of the side chain of engineered cysteines tethered to Fe-EDTA is about \( d = 14 \text{ Å} \). 80% of the distances lie within \( d \pm 50\% \) which is compatible with the diffusion rate of the generated hydroxyl radicals.
Figure 2.1 Sites of mutagenesis to introduce Cys residues and subsequent tethering of EDTA-Fe. (A) Comparison of the sequences of the protein subunits of RNase P from *E. coli* and *B. subtilis*. The residues marked by an asterisk indicate sites of Cys replacement and modification with EPD-Fe in the protein subunit of *E. coli* RNase P. The positions of the respective secondary structural elements are also indicated. (B) Alignment of the amino acid sequence around helix α2 in the protein subunit of RNase P from 12 different bacteria. (C) Tertiary structure of the protein subunit of RNase P from *B. subtilis* (Stams et al., 1998). The program RASMOL was used to depict the α-carbon backbone of the protein structure as a ribbon. The various secondary structural elements are depicted using the same color scheme as in Panel A. The positions where Cys residues were introduced and modified with EPD-Fe are labeled. Numbering is based on the sequence of the protein subunit of *E. coli* RNase P. (D) The chemical reaction for covalent attachment of EDTA-Fe to the single Cys residue engineered in a protein.
Figure 2.2 OH\(^{-}\)-mediated cleavages in M1 RNA in the presence of unmodified and EDTA-Fe-modified derivatives of C5 protein. Selected regions of the autoradiograms in panels A, B, C and D were analyzed with a phosphorimager (Molecular Dynamics) to illustrate the cleavages in M1 RNA promoted by C5 Cys52 EDTA-Fe (panel E), C5 Cys59-EDTA-Fe (panel F), C5 Cys62-EDTA-Fe (panel G), and C5 Cys70-EDTA-Fe (panel H). The scans of the footprinting reactions performed with the unmodified and EDTA-Fe-modified proteins are shown in blue and red, respectively. The numbers indicate the position of the nucleotides in M1 RNA. The corresponding sequence in M1 RNA is also provided. Since the peak heights are in arbitrary units, no scale has been provided for the y-axis. The cleavage sites were assigned using reference standards as described earlier (Biswas et al, 2000).
Figure 2.3 Summary of footprinting data. (A) The OH·-mediated cleavages in M1 RNA promoted by C5 Cys16-, Cys54- and Cys66-EDTA-Fe are denoted by red circles, blue triangles, and green squares, respectively (Biswas et al., 2000). (B) The OH·-mediated cleavages in M1 RNA promoted by C5 Cys52-, Cys59-, Cys62- and Cys70-EDTA-Fe are denoted by red circles, brown stars, green squares and light blue triangles, respectively. The size of the symbols corresponds to the intensity of the hydroxyl radical-mediated cleavages. The secondary structure representation and the numbering of helices of M1 RNA are as described by Massire et al. (1998).
Figure 2.4 Crosslinking of the E. coli RNase P holoenzyme-substrate complex. (A) Secondary structure of ptRNAPhe used in this study. The arrow indicates the cleavage site of RNase P. (B) Separation of free and crosslinked ptRNAPhe using SDS-PAGE. All lanes correspond to reactions either with or without an RNase P holoenzyme that have been subjected to UV irradiation at 312 nm in the presence of radiolabeled ptRNAPhe. The holoenzymes were reconstituted using M1 RNA and a single Cys-substituted mutant derivative of C5 protein either unmodified or modified with azidophenacyl (AzP) bromide. (C) Primer extension analysis to map the position in ptRNAPhe which is crosslinked to Cys50 in C5 protein (see text for details).
Figure 2.5 Model of the bacterial RNase P holoenzyme. (A) A stereoview of the three-dimensional model of *E. coli* RNase P built using the footprinting data obtained in this study. (B) An illustration depicting the amino acid positions where EDTA-Fe was conjugated and the corresponding OH⁻-mediated cleavages in M1 RNA. The side chains of the amino acid residues and the spheres representing their respective cleavage sites in M1 RNA are colored identically to facilitate an appreciation of how docking of C5 protein on M1 RNA was achieved. (C) Large white spheres indicate the positions of the invariant nucleotides among all the RNA subunits of bacterial RNase P (Massire et al., 1998). To orient the reader, secondary structural elements in the RNA and protein subunits of *E. coli* RNase P are labeled in the various panels.
Figure 2.6 Model of the bacterial RNase P holoenzyme-substrate complex. (A) A stereo view of the three-dimensional model of the *E. coli* RNase P holoenzyme-ptRNA complex. The color scheme of M1 RNA is the same as in Figure 5. C5 protein and the ptRNA substrate are depicted in cyan and silvery white color, respectively. (B) A stereo view of a magnified section of panel A depicting the base-pairing between the 3’-ACCA terminus of the ptRNA substrate with G292 and G293 in P15 of M1 RNA, and the proximity of nucleotide –4 in the ptRNA leader sequence and amino acid residue 50 in C5 protein (both shown in ball and stick format). This distance constraint is based on the crosslinking studies (see text for details).
CHAPTER 3

FUNCTIONAL RECONSTITUTION AND CHARACTERIZATION OF RNASE P FROM *PYROCOCCUS FURIOSUS*: INSIGHTS INTO THE TRANSITION OF A CATALYTIC RNA TO AN RNP COMPLEX

3.1 Introduction

Ribonuclease P (RNase P) is an ancient and essential endoribonuclease that catalyzes the 5'-end maturation of tRNAs in all three domains of life (Guerrier-Takada et al, 1983; Kirsebom & Vioque, 1995; Frank & Pace, 1998; Hall & Brown, 2001; Jarrous & Altman, 2001; Gopalan et al, 2002; Xiao et al, 2002; Hartmann & Hartmann, 2003; Hsieh et al, 2004). While RNase P in all living organisms contains one essential RNA subunit, the number of protein cofactors/subunits varies: at least one in bacterial (Kole & Altman, 1981), four in archaeal (Hall & Brown, 2002), and nine in eukaryal (nuclear) RNase P (Chamberlain et al, 1998; Jarrous, 2002). The basis for this variability, which has implications for macromolecular evolution, remains to be elucidated.

The RNA component alone from bacterial RNase P is catalytically active under *in vitro* conditions of high ionic strength in the presence of a divalent ion such as Mg$^{2+}$ (Guerrier-Takada et al, 1983); however, the protein cofactor is essential for RNase P
function *in vivo* due to its pleiotropic effects on RNA structure and catalysis, such as global stabilization of tertiary structure, improved substrate recognition, higher affinity for Mg\(^{2+}\), and faster precursor tRNA (ptRNA) cleavage (Reich et al, 1988; Tallsjo & Krsebom, 1993; Crary et al, 1998; Niranjanakumari et al, 1998; Kurz et al, 2002; Hsieh et al, 2004; Buck et al, 2005). Although phylogenetic sequence comparison analysis established that the RNA subunits of archaeal and eukaryal RNase P retain the same catalytic core as the bacterial ribozyme, most archaeal and all eukaryal RNase P RNAs are inactive in the absence of their cognate protein subunits under various conditions when tested *in vitro* (Haas et al, 1996; Frank & Pace, 1998; Pannucci et al, 1999; Harris et al, 2001; Xiao et al, 2002). Genetic and biochemical studies established the association of yeast/human nuclear RNase P activity with at least nine protein subunits (seven of which are homologs) (Chamberlain et al, 1998; Jarrous, 2002). By using polyclonal antisera generated against four *Methanothermobacter thermoautotrophicus* (*Mth*) polypeptides, which were identified by database mining as homologs of four out of the seven conserved yeast/human RNase P protein subunits, RNase P activity could be immunoprecipitated from a partially purified *Mth* RNase P preparation (Hall & Brown, 2002). None of the archaeal or eukaryal RNase P protein subunits exhibits notable sequence homology to the sole protein associated with bacterial RNase P (Hartmann & Hartmann, 2003).

The increasing protein content in archaeal/eukaryal RNase P has raised the intriguing question of whether the catalytic role of the RNA subunit in bacteria has been usurped by some of the multiple protein subunits in archaea/eukarya. An essential first step to address this query is to reconstitute *in vitro* archaeal/eukaryal RNase P.
holoenzymes and perform biochemical studies such as those performed with bacterial RNase P. Indeed, reconstitution using some or all recombinant components has been recently reported for archaeal \textit{Pyrococcus horikoshii} (Pho; Kouzuma et al, 2003) and \textit{Mth} (Boomershine et al, 2003) and eukaryal (human) RNase P (Mann et al, 2003). These earlier studies represented a significant advance in suggesting the feasibility of \textit{in vitro} reconstitution of an RNase P complex hitherto proven to be intractable. In these cases, however, the reconstituted RNase P holoenzymes displayed weak activity (compared to their native versions or holoenzymes from other species) and only under single-turnover conditions. Moreover, notably absent were core platform(s) representing partial assemblages that one might typically expect as intermediates in a hierarchical biogenesis pathway. Therefore, it was clear that further exploration of different sample preparation approaches or additional assay conditions were needed to establish robust \textit{in vitro} reconstitution under multiple turnover conditions necessary meaningful structure-function analyses. Towards this objective, we report here the successful functional reconstitution of \textit{Pfu} RNase P and the resulting insights into its functional organization and assembly.

3.2 Materials and methods

3.2.1 Cloning of the RNA and protein subunits of \textit{Pfu} RNase P.

Database mining of the \textit{Pfu} genome confirmed the presence of an RNA and four protein subunits, which share homology to their well established yeast/human counterparts (Table 3.1).
The gene encoding \textit{Pfu} RPR was obtained by PCR with \textit{Pfu} genomic DNA as the template and the gene-specific primers listed in Table S2. The PCR product containing the RPR gene was digested with \textit{Eco}RI (whose recognition site was included in the reverse primer) and then subcloned into pBT7 (Tsai et al, 2002) that had been digested with \textit{Stu}I (which generates a blunt end) and \textit{Eco}RI. The resulting plasmid named pBT7-\textit{Pfu} RPR has the \textit{Pfu} RPR gene placed under the control of a T7 RNA polymerase promoter. \textit{Pfu} RPR can be generated using either \textit{Eco}RI- or \textit{Sty}I-linearized pBT7-\textit{Pfu} RPR as template in an \textit{in vitro} transcription reaction.

The genes encoding Pop5, Rpp21, Rpp29 and Rpp30 were amplified by PCR with \textit{Pfu} genomic DNA as the template and the gene-specific primers listed in Table 3.2. The PCR products were then digested with \textit{Eco}RI, whose recognition site was included in the different reverse primers. All four protein-encoding ORFs were cloned into pET-33b that had been digested with \textit{Nco}I, filled in with Klenow, and subsequently digested with \textit{Eco}RI. This cloning approach ensured that no additional amino acid residues were introduced in the four ORFs.

Automated DNA sequencing was used to establish the authenticity of all the clones obtained using the procedures described above.

\textbf{3.2.2 Generation of RPRs using \textit{in vitro} transcription.}

pBT7-\textit{Pfu} RPR and was linearized with \textit{Eco}RI and used as templates for T7 RNA polymerase-mediated run-off transcription using established protocols (Vioque et al, 1988; Tsai et al, 2002). The RNA transcripts thus generated were subsequently purified...
by either dialysis or a QuickSpin (gel filtration) column procedure (Vioque et al, 1988; Tsai et al, 2002).

3.2.3 Protein overexpression.

*Escherichia coli* BL21(DE3) Rosetta cells (Novagen) were transformed with pET-33b plasmids containing the gene for one of the four protein subunits of *Pfu* RNase P. Cells were grown at 37 °C in LB media containing 35 µg/ml kanamycin and 35 µg/ml chloramphenicol and induced with 2 mM isopropyl-β-D-thiogalactoside (IPTG) at OD$_{600}$ ~ 0.6. For Rpp21 overexpression, 50 µM ZnCl$_2$ was also added at the time of induction as the protein was suspected to harbor a Zn$^{2+}$-binding motif (based on its sequence), a premise that has been borne out by recent structural studies (Boomershine, 2005; Kakuta et al, 2005). All four *Pfu* Rpps overexpressed well and were purified using cation exchange chromatography with the exception of Rpp21 (Fig. 3.1; see below).

3.2.4 Purification of protein subunits of *Pfu* RNase P

3.2.4.1 *Pfu* Pop5

Cells were resuspended in denaturing buffer [25 mM Tris-HCl (pH 7), 7 M urea, 10 mM DTT, 1 mM EDTA, 0.1 mM PMSF], sonicated, and centrifuged for 15 min at 12,000 g. The supernatant was passed through a 0.22-µm filter and loaded on a 5-ml HiTrap SP-Sepharose column (Amersham Pharmacia). Pop5 was re-folded on the column using a decreasing (7 to 0 M) urea gradient and subsequently eluted using an increasing salt gradient. Pop5 typically eluted around 1.5 M NaCl.
3.2.4.2 *Pfu Rpp29*

Cells were resuspended in non-denaturing buffer [25 mM Tris-HCl (pH 7.5), 1 mM DTT, 0.1 mM PMSF, and 1 mM EDTA], sonicated, and centrifuged for 15 min at 12,000 g. The pellet was then resuspended in denaturing buffer, sonicated, and centrifuged for 15 min at 12,000 g. The supernatant was filtered using a 0.22-µm filter and loaded on a 5-ml HiTrap SP-Sepharose column (Amersham Pharmacia). Rpp29 was then re-folded and eluted as described above for Pop5. Rpp29 typically eluted around 1 M NaCl.

3.2.4.3 *Pfu Rpp30*

We employed two cation-exchange chromatography steps using the same matrix (SP-Sepharose) but under successive denaturing and native conditions. We first prepared the sample in a manner similar to *Pfu Pop5* (see above) up to the step of loading the SP-Sepharose column. Rpp30 was then eluted using an increasing salt gradient under denaturing conditions (i.e., 7 M urea). Peak fractions containing partially purified Rpp30 were pooled, dialyzed to remove salt, re-loaded on an SP-Sepharose column, re-folded using a decreasing (7 to 0 M) urea gradient and eluted using an increasing salt gradient. Rpp30 typically eluted around 550 mM NaCl.

3.2.4.4 *Pfu Rpp21*

Cells were resuspended in non-denaturing buffer, sonicated, and centrifuged for 15 min at 12,000 g. The pellet, containing *Pfu* Rpp21, was washed successively with the non-denaturing buffer supplemented with 3 M and 7 M urea to help solubilize proteins other than Rpp21. The final wash was centrifuged for 20 min at 14,500 g. The pellet,
containing *Pfu* Rpp21, was resuspended in a stronger denaturing buffer [25 mM Tris-HCl (pH 7.5), 0.1 mM PMSF, 5 mM DTT, and 6 M guanidine hydrochloride], sonicated, and centrifuged for 20 min at 14,500 g. The supernatant was filtered using a 0.22-µm filter and applied on a C4 reversed-phase HPLC column. Elution was performed using a linear gradient of CH₃CN in 0.1% trifluoroacetic acid (TFA). Rpp21 typically eluted at 40% CH₃CN. The peak fractions were pooled, lyophilized, and resuspended in a highly acidic buffer containing 50 mM Tris-HCl (pH 2.5), 10 mM MgCl₂, 800 mM NH₄OAc, 1 mM DTT, 300 μM ZnCl₂. The pH of the solution was slowly raised to 7.5 by adding 10 M NaOH (modified from Boomershine, 2005).

The A₂₈₀ values for the different *Pfu* Rpps were measured and the concentrations calculated using their respective extinction coefficients. Purified *Pfu* Rpps were stored at room temperature and dialyzed against 1X assay [50 mM Tris-HCl (pH 8), 800 mM NH₄OAc and 30 mM MgCl₂] buffer prior to use.

### 3.2.5 *Pfu* RNase P reconstitution and assay.

*Pfu* RPR transcripts (in water) were folded by incubating for 50 min at 50 °C, 10 min at 37 °C, and, then for 30 min at 37 °C after addition of 1/10th volume 10X assay buffer. Reconstitution experiments were initiated by pre-incubating folded *Pfu* RPR with either all or a subset of four *Pfu* Rpps in assay buffer [50 mM Tris-HCl (pH 8), 800 mM NH₄OAc and varying amounts of MgCl₂] for 5 min at 37 °C followed by 10 min at 55 °C. The activity was subsequently assayed by adding 500 nM *E. coli* ptRNA₄Tyr, a trace amount of which was internally labeled with [α-³²P] GTP. After defined time intervals at
55 °C, the reactions were quenched with urea-phenol dye [8 M urea, 0.04% (w/v) bromophenol blue, 0.04% (w/v) xylene cyanol, 0.8 mM EDTA, 20%(v/v) phenol].

It was important that comparative analyses with partially and fully reconstituted \textit{Pfu} RNase P complexes were performed under identical assay conditions. However, this meant that the reconstitution of \textit{Pfu} RPR with four Rpps would yield a complex with much higher activity than the partial complexes. A highly active enzyme necessitates assay time-points of short intervals (e.g., 5 s) if the amount of substrate cleaved is to be restricted to <30%. In these instances, reactions were first terminated by quickly plunging the assay reaction tubes into liquid nitrogen and the urea-phenol dye was added subsequently.

Uncleaved precursor tRNA\textsubscript{Tyr} and its products (generated by RNase P) were separated using 8% (w/v) polyacrylamide/7 M urea gel electrophoresis and visualized by autoradiography. The extent of reaction was quantitated by phosphorimager analysis of the polyacrylamide gels.

### 3.3 Results

#### 3.3.1 Nomenclature

The \textit{Pfu} RNase P RNA subunit is referred here as \textit{Pfu} RPR. The RNase P protein subunits are named with the prefix Rpp together the molecular weight of the homologs associated with human RNase P (\textit{Pfu} Rpp21, Rpp29 and Rpp30). \textit{Pfu} Pop5 is so named based on nomenclature that was developed for the \textit{pop} mutants in yeast, i.e. those with defects in processing of precursor RNAs.
3.3.2 Rationale

We chose to study Pfu RNase P for three main reasons. First, the high sequence conservation between protein subunits of archaeal and eukaryal RNase P makes it likely that insights on how Pfu RPPs modulate Pfu RPR catalysis might be applicable to eukaryal RNase P, whose in vitro reconstitution has proven difficult. Second, with fewer (predicted) and smaller proteins compared to its eukaryal relatives, Pfu RNase P is a simpler ribonucleoprotein (RNP) complex (one RNA and four proteins) to study although its Rpps might represent the core proteins in eukaryal RNase P. Lastly, the Pfu Rpps are highly positively charged and thermostable, features that we expected to facilitate their purification and high-resolution structural studies.

An important caveat needs to be emphasized at the outset. Although extensive biochemical characterization of partially purified native RNase P from Mth and Methanococcus jannaschii (Mja) has been reported (Andrews et al, 2001), the protein subunit composition in these native RNase P preparations has not been established. Searches of archaeal genome sequences and results from immunoprecipitation experiments (Hall & Brown, 2002) support the presence of a single RNA and at least four protein subunits in archaeal RNase P. However, the possibility of additional protein subunit(s) in archaeal RNase P, not predicted by database mining efforts undertaken to identify sequences resembling eukaryal homologs, will remain until proven otherwise by elucidation of the subunit composition of a native holoenzyme. Therefore, the most logical in vitro reconstitution components for archaeal RNase P currently comprise its sole RNA subunit and four protein subunits.
3.3.3 Purification of the recombinant subunits of \textit{Pfu} RNase P.

\textit{Pfu} RPR was generated using run-off \textit{in vitro} transcription and purified to homogeneity using standard protocols (Vioque et al, 1988; Tsai et al, 2002). The protein subunits were overexpressed in \textit{E. coli} BL21(DE3) Rosetta cells and purified to homogeneity using cation-exchange or reversed-phase chromatography (Fig. 3.1). Electrospray ionization mass spectrometry was used to verify the integrity and identity of the four purified protein subunits of \textit{Pfu} RNase P. The molecular masses observed are in excellent agreement with those predicted by the amino acid sequences (Table 3.1).

3.3.4 \textit{In vitro} reconstitution of \textit{Pfu} RNase P.

To examine whether the putative recombinant protein subunits of \textit{Pfu} RNase P could promote \textit{Pfu} RPR function \textit{in vitro}, we employed a standard ptRNA-processing assay (Guerrier-Takada et al, 1983; Vioque et al, 1988). \textit{Pfu} RPR reconstituted with its four Rpps exhibits robust processing of \textit{E. coli} ptRNA\textsubscript{Tyr} under multiple turnover conditions, an important feature of a true biological catalyst (Fig. 3.2, lanes 4 and 19). The optimal substrate concentration to be used in the assays was obtained by testing different concentrations of ptRNA\textsubscript{Tyr} with reconstituted \textit{Pfu} RNase P and determining which substrate concentration resulted in the highest turnover number. Such an experiment revealed that 500 nM of ptRNA\textsubscript{Tyr} was optimal (data not shown). Using optimal conditions for reconstitution of \textit{Pfu} RNase P activity and assuming that 100% of \textit{Pfu} RPR in the reaction assembled with the protein subunits, we obtain turnover numbers up to 15 min\textsuperscript{-1}. This number decreases as the time between the purification of proteins and their use in the assay increases.
Some control experiments merit mention. First, to establish that the ptRNA-processing activity requires both the RNA and protein components of the *Pfu* RNase P complex, we performed assays with *Pfu* RPR alone (Fig. 3.2, lane 1) or just the four *Pfu* Rpps (Fig. 3.2, lane 3). Neither reaction showed any detectable activity. Second, since all four *Pfu* Rpps were purified subsequent to overexpression in *E. coli*, the remote possibility existed that the activity observed in our reconstitution studies was due to a low (undetectable by SDS-PAGE) level of contamination of C5 protein (the protein cofactor of *E. coli* RNase P) which somehow heterologously reconstituted with *Pfu* RPR. To address this caveat, we reconstituted *Pfu* RPR with highly purified recombinant C5 protein and found no ptRNA-processing activity (Fig. 3.2, lane 2). Finally, to ensure that the ptRNA-processing activity obtained with the reconstituted *Pfu* RNase P was indeed a *bona fide* RNase P activity, and not due to non-specific cleavage, we mapped the precise cleavage site in the ptRNA substrate subsequent to processing by either the *in vitro* reconstituted *Pfu* RNase P or a partially purified preparation of native *Pfu* RNase P (data not shown). Primer extension analysis of the cleaved product revealed that both reconstituted and native *Pfu* RNase P cleaved ptRNA<sup>Tyr</sup> at the expected position i.e., between the −1 and +1 position, the first nucleotide of the mature tRNA (data not shown). Taken together, these results confirm that the enzymatic activity exhibited by the reconstituted *Pfu* RNase P is *bona fide* and attributable to functional reconstitution of its RNA and protein subunits.
3.3.5 Reconstitution of partially and fully reconstituted *Pfu* RNase P.

To examine whether all four proteins are needed for reconstituting a robust enzyme capable of multiple turnover, we assayed the RNA subunit with partial suites of the four protein subunits under multiple-turnover conditions (Fig. 3.2, lanes 5-18). No RNase P activity was observed when only one of the four *Pfu* Rpps is added to *Pfu* RPR. Among the six possible two-protein combinations, only two were catalytically active: Pop5+Rpp30 and Rpp21+Rpp29 (Fig. 3.2, lanes 11 and 12). As expected from the pair-wise combination results, all four combinations of three protein subunits display activity when reconstituted with *Pfu* RPR (Fig. 3.2, lanes 15-18). These data suggest that the *Pfu* RNase P holoenzyme can be built using two initiating RNP complexes as the minimal functional cores: *Pfu* RPR+Pop5+Rpp30 and *Pfu* RPR+Rpp21+Rpp29. However, *Pfu* RPR+Pop5+Rpp30 is considerably more active than *Pfu* RPR+Rpp21+Rpp29 (Fig. 3.2, lanes 11 and 12).

To determine the contribution of individual protein subunits to *Pfu* RPR catalysis, we compared the initial velocity of ptRNA processing exhibited by *Pfu* RPR in the presence of different two- and three-protein combinations (i.e., partially reconstituted *Pfu* RNase P). Prior to embarking on such a comparative study, it was important to establish the optimal conditions for assaying these different RNP complexes.

Since RNase P requires a divalent metal ion (generally Mg$^{2+}$) for catalysis (Guerrier-Takada et al, 1983), we entertained the idea that different combinations of protein subunits with *Pfu* RPR might have a different optimal Mg$^{2+}$ concentration for eliciting the maximum activity, a premise not considered in earlier studies (Boomershine et al, 2003; Kouzuma et al, 2003). While *Pfu* RPR is reported to be weakly active at 300
mM $\text{Mg}^{2+}$ and a high concentration of monovalent ions (4 M $\text{NH}_4^+$; Pannucci et al, 1999), the $\text{Mg}^{2+}$ requirement for optimal catalytic activity is drastically decreased to 120 and 30 mM when $Pfu$ RPR is reconstituted with two and four protein subunits, respectively (Figs. 3.3A and 3.3B). An intermediate value is also observed when addition of Pop5 to $Pfu$ RPR+Rpp21+Rpp29 results in a decrease from 120 to 40 mM $\text{Mg}^{2+}$ (Fig. 3.3B). Addition of two or more protein subunits to $Pfu$ RPR clearly reduces its $\text{Mg}^{2+}$ requirement for optimal activity. Incidentally, the 30 mM $\text{Mg}^{2+}$ optimum determined for $Pfu$ RPR reconstituted with the four $Pfu$ Rpps is identical to that reported for the native $Mja$ RNase P holoenzyme (Andrews et al, 2001). However, it is important to appreciate that the optimal $\text{Mg}^{2+}$ might vary based on the source of enzyme and the assay conditions.

Since we do not know the $K_D$ values for the binding of the protein subunits to the RNA moiety, we had to determine if a stoichiometric excess of recombinant protein subunits was required to promote near-complete assembly of the RNA subunit to form the different RNP complexes. By titrating different amounts of the protein subunits to a constant amount of RNA subunit (50 nM) and assaying for activity under optimal $\text{Mg}^{2+}$ concentrations, we observed that a 10:1 ratio of protein:RNA was optimal for most of the partial holoenzymes complexes and that a 2.5:1 ratio was adequate for the fully reconstituted $Pfu$ RNase P (Fig. 3.3C and 3.3D). While this result indicates cooperativity among the subunits during holoenzyme assembly, it does not comment on stoichiometry.

Once these preliminary experiments were completed, the initial velocities were calculated under optimal conditions for complexes nucleated from two minimal core
initiating units: \( Pfu \) RPR+Pop5+Rpp30 or \( Pfu \) RPR+Rpp21+Rpp29 (Fig. 3.3E). The y-axis in Figure 3.3E is plotted in logarithmic scale to emphasize the dramatic changes in activity that occur upon sequential addition of the protein subunits. The turnover numbers are summarized in Figure 3F. Of the two minimal functional core complexes, the initial velocity of \( Pfu \) RPR+Pop5+Rpp30 is 11-fold greater than \( Pfu \) RPR+Rpp21+Rpp29. Addition of neither Rpp21 nor Rpp29 singly to \( Pfu \) RPR+Pop5+Rpp30 has any appreciable effect on activity. However, adding both Rpp21 and Rpp29 together to \( Pfu \) RPR+Pop5+Rpp30 results in a 23-fold increase in initial velocity. With \( Pfu \) RPR+Rpp21+Rpp29 as the core RNP complex, addition of Rpp30 had no effect while addition of Pop5 elicited a 3-fold increase. When both Rpp30 and Pop5 are added to \( Pfu \) RPR+Rpp21+Rpp29, a 250-fold increase in activity was observed.

3.4 Discussion

3.4.1 In vitro reconstitution of \( Pfu \) RNase P.

In this study, we have optimized the conditions for reconstituting the \( Pfu \) RNase P holoenzyme from its constituent RNA and four protein subunits and demonstrated that this enzyme can perform robust multiple turnover, an important consideration for any biological catalyst. Although Kimura and coworkers reported reconstitution of the \( P. \) horikoshii RNase P holoenzyme (Kouzuma et al, 2003), a first for any archaeal or eukaryal RNase P, it was under single turnover conditions that do not permit structure-function relationships to be dissected in any meaningful fashion. The \( k_{cat} \) of our in vitro reconstituted \( Pfu \) RNase P enzyme under multiple turnover conditions
(~10-15 min\(^{-1}\)) is only a few fold lower than those reported for the native RNase P
holoenzymes purified from archaeal sources (~34 to 52 min\(^{-1}\); Andrews et al, 2001) and
yeast (~72 min\(^{-1}\); Xiao et al, 2005).

In addition to considerations such as assay buffer and ptRNA substrate used, there
are various reasons (including a few listed below) why an additional three- to four-fold
increase in turnover number of \(Pfu\) RNase P might yet be realizable. First, the turnover
number that we have calculated could be an underestimate if not all of the \(Pfu\) RPR
assembles into a holoenzyme complex. We believe that this limitation could be
addressed by reducing the variables during reconstitution such as the use of binary
complexes of interacting protein subunits that were purified as such. Second, as pointed
out previously, it is conceivable that additional protein subunit(s) might be present in
archaeal RNase P and that the four we have used represents an incomplete set. We will
soon attempt to identify the protein subunit composition of an archaeal native RNase P
holoenzyme. Lastly, the recombinant proteins purified after overexpression in \(E. coli\)
may not be completely re-folded (mirroring their native environment) or have the
post-translational modifications that normally occur in archaia. This latter caveat is not
so easily dealt with until further advances in our understanding of native archaeal
RNase P.

### 3.4.2 Hierarchical assembly of \(Pfu\) RNase P

The stringent assay conditions used for reconstitution of \(Pfu\) RNase P activity have
allowed us to focus on questions about its assembly. A comparison of the optimal
activities of partially reconstituted \(Pfu\) RNase P revealed that \(Pfu\) RPR+Rpp21+Rpp29
and RPR+Pop5+Rpp30 constitute minimal functional cores and that a dramatic increase in the activity of reconstituted \textit{Pfu} RNase P results upon addition of the remaining protein subunits to these core complexes. It is important to note that \textit{Pfu} RPR+Pop5+Rpp30 is more active than \textit{Pfu} RPR+Rpp21+Rpp29 (Fig. 3.3). Recently, we have confirmed this pattern of incremental activity with \textit{Mth} RNase P as well (data not shown). Although these partially active complexes were not identified in two earlier reports on archaeal RNase P (including ours on \textit{Mth}; Boomershine et al, 2003), most likely due to suboptimal Mg\textsuperscript{2+} concentrations in the assay, the two core RNP complexes that we have now identified are expected functional attributes based on results from previous genetic and biochemical studies. First, yeast two-hybrid analyses revealed protein-protein interactions between Pop5 and Rpp30 and between Rpp21 and Rpp29 in archaeal (\textit{Mth} and \textit{Pho}) and eukaryal (human and yeast) RNase P (Jiang & Altman, 2001; Houser-Scott et al, 2002; Hall & Brown, 2004; Kifusa et al, 2005). Second, yeast three-hybrid analyses indicated that human Rpp21, Rpp29 and Rpp30 individually interacted with the human RPR while yeast Rpp29 similarly interacts with its cognate RPR (Jiang et al, 2001). Third, when human RPR was reconstituted with human Rpp21 and Rpp29, weak activity was observed (Mann et al, 2003). Lastly, NMR spectroscopic studies of the \textit{Pfu} Rpp21+Rpp29 and Pop5+Rpp30 complexes revealed that there were significant chemical-shift perturbations in the HSQC spectrum of each member of each pair upon addition of its partner, indicating strong macromolecular interactions within the pairs (Boomershine, 2005; Wilson et al, 2005).

Bacterial RPRs, which are active in the absence of their cognate protein subunits, have been classified into two distinct structural types, A and B (Hall & Brown 2001).
More recently, archaeal RPRs were also categorized into two types, A and M (the latter primarily from Methanococci; Harris et al, 2001). Archaeal type A RPRs, exemplified by Mth and Pfu, are structurally and functionally similar to the ancestral bacterial type A RPR; they can exhibit catalytic activity in the absence of their cognate protein subunits in 4 M NH$_4$ and 300 mM Mg$^{2+}$ (Pannucci et al, 1999). Archaeal type M RPRs, typified by Mja, differ from their type A counterparts in that they are inactive in the absence of protein subunits presumably due to the absence of critical structural elements necessary for substrate recognition (Pannucci et al, 1999; Harris et al, 2001). It will be of interest to investigate if the minimal functional cores and the increase in activity observed during the hierarchical assembly of Mth and Pfu RNase P (type A) will also be seen in Mja RNase P (type M).

3.4.3 Cooperativity during assembly of Pfu RNase P.

We used a gel-shift assay to assess the ability of each of the four Pfu Rpps to bind Pfu RPR either individually or as binary complexes. Consistent with the activity assays where no single protein subunit could confer even weak activity on the RNA (Fig. 3.2), we failed to detect specific binding between individual protein subunits and the RNA moiety (data not shown). Although we could detect some mobility-retarded species consistent with the presence of RNA bound to binary complexes, non-specific large aggregates have prevented us from obtaining reliable K$_D$ values. In the absence of this information, we had to rely on protein titration experiments to determine the protein:RNA ratios that would permit the maximum RNP complex formation for the various Pfu RPR and Rpp combinations. Even in the absence of detailed binding studies,
these titration experiments indicate cooperativity among the subunits during assembly of *Pfu* RNase P (Fig. 3.3C and 3.3D).

As discussed above, results from biochemical and genetic studies indicate that heterodimers composed of Rpp30+Pop5 and Rpp21+Rpp29 can form in the absence of the cognate RNA. Direct evidence for the order of events during *Pfu* RNase P assembly is still lacking. However, the observations that (i) *Pfu* RPR displays basal activity with two non-overlapping pairs of protein subunits but not the individual subunits (Fig. 3.2), and (ii) the addition of a third protein subunit to a pre-formed ternary complex of one RNA and two protein subunits has little effect on activity (with one exception; Fig. 3.3F), argue in favor of interacting protein subunit pairs being the RNA activators regardless of whether heterodimer formation occurs prior to RNA binding or on the RNA scaffold. Our results also indicate that two different ternary complexes made up of an interacting pair of protein subunits bound to the RNA could serve as a nucleating intermediate en route to the complete complex of one RNA and four proteins. Therefore, assembly of *Pfu* RNase P is not an all or none situation where the RNA has to simultaneously bind all four protein subunits to form the holoenzyme.

The cooperativity among protein subunits would imply an increase in the affinity of a binary complex (e.g., Rpp30+Pop5) for an RNA already loaded with the other binary complex (i.e., Rpp21+Rpp29). Although this remains to be established, parallels in this regard have been noted in other RNP complexes such as the ribosome and telomerase (Mizushima & Nomura, 1970; Recht & Williamson, 2004; Prathapam et al, 2005). In the small subunit of the bacterial ribosome, the central domain of 16S rRNA offers a platform wherein a core RNP consisting of six ribosomal proteins is formed.
During assembly, the rRNA interacts with a primary protein (S15) which enhances the affinity of the rRNA for two secondary proteins (S6+S18) that bind as a heterodimer (Recht & Williamson, 2004). The presence of the p65 protein in the *Tetrahymena thermophila* telomerase holoenzyme enhances the affinity of the reverse transcriptase for the telomerase RNA by nearly 30-fold (Prathapam et al, 2005).

### 3.4.4 Role of protein cofactors/subunits in *Pfu* RNase P catalysis.

Studies on bacterial RNase P have revealed that its sole protein subunit plays three roles in aiding the RNA catalyst: (i) enhancing affinity of the catalytic RNA moiety for the ptRNA substrate (in large part due to a binding pocket that accommodates the ptRNA leader), (ii) increasing affinity for Mg$^{2+}$, and (iii) increasing the rate of cleavage (Reich et al, 1988; Tallsjo & Kirsebom, 1993; Crary et al, 1998; Niranjanakumari et al, 1998; Kurz & Fierke, 2002; Hsieh et al, 2004; Buck et al, 2005a). It is premature to speculate on which archaeal RNase P protein subunits, either individually or in pairs, might specifically undertake these roles, since the precise contributions of each Rpp to catalysis by *Pfu* RPR remains to be dissected by kinetic and thermodynamic studies. Nevertheless, our results warrant some remarks.

*Pfu* Pop5+Rpp30 elicits a ~11-fold higher activity with *Pfu* RPR than Rpp21+Rpp29, suggesting their importance in *Pfu* RNase P catalysis. While ptRNA processing might proceed with a partial suite of protein subunits or even the RNA alone under certain conditions, an examination of the relative activities of the various combinations of *Pfu* RPR and Rpps (Fig. 3.3F) reveals that no single protein subunit is
dispensable for robust activity, a result consistent with the finding that all protein subunits are essential for viability in yeast (Chamberlain et al, 1998).

In the case of \textit{Mth} RNase P, the observation that the affinity of the substrate for the native holoenzyme might be at least 1000-fold higher than that observed with \textit{Mth} RPR alone led to the important conclusion that the protein cofactors must play a significant role in substrate recognition akin to the bacterial RNase P scenario (Pannucci et al, 1999; Andrews et al, 2001). In this regard, it is interesting that Wilson et al. (2005) have proposed that Pop5 might be involved in ptRNA recognition due to its striking structural similarity to the bacterial RNase P protein cofactor. Further studies are required to demonstrate functional equivalence in these two proteins that share little primary sequence homology.

Metal ions play vital roles in RNA structure and catalysis. In a postulated reaction mechanism, a hydrated Mg\textsuperscript{2+} ion is believed to play a direct role in RNase P catalysis by generating the hydroxide nucleophile which performs an S\textsubscript{N}2 in-line attack on the phosphodiester linkage in the ptRNA substrate. \textit{Pfu} RPR alone is active in 4 M NH\textsubscript{4}\textsuperscript{+} and 300 mM Mg\textsuperscript{2+} (Pannucci et al, 1999). The two-protein minimal functional cores (\textit{Pfu} RPR+Rpp21+Rpp29 and \textit{Pfu} RPR+Rpp30+Pop5) as well as three-protein complexes (a minimal core plus an additional protein) function optimally at 120 mM with one exception (40 mM for \textit{Pfu} RPR+Pop5+Rpp21+Rpp29; Fig. 3.3B). \textit{Pfu} RNase P reconstituted with all four protein subunits exhibits maximum activity at 30 mM Mg\textsuperscript{2+}. Clearly, protein-mediated increase in the affinity of the RNA moiety for Mg\textsuperscript{2+}, a recurring theme in protein-aided RNA catalysis in RNP complexes (Solem et al, 2002),
is being employed effectively by archaeal RNase P in a manner reminiscent of its bacterial counterpart (Kurz & Fierke, 2002; Hsieh et al, 2004).

3.4.5 Summary.

A premise of the RNA world hypothesis is that protein cofactors were recruited either to take over some functions of or to render RNA catalysts (from a primordial stage) more efficient and versatile in the extant cellular milieu (Joyce, 2002). Our findings on archaeal RNase P illustrate and support the possibility of such an evolutionary transition. Specifically, we found that archaeal RNase P RNA can functionally reconstitute with two different pairs of protein cofactors suggesting that mobilizing even two of the four extant protein subunits would have drastically improved the efficiency of the ancestral RNA enzyme. Also, this raises the possibility that recruitment of protein cofactors could have proceeded in a chronological fashion via intermediates that displayed sequential improvements in catalytic efficiency.

RNase P protein cofactor(s) in the bacterial and archaeal lineages appear to have converged on similar solutions to facilitate RNA catalysis (such as increased substrate affinity and lowering the concentration of optimal Mg$^{2+}$). What still remains unresolved is the reason for increased protein complexity in archaeal/eukaryal RNase P compared to the bacterial relative. The recent determination of the tertiary structures of all four protein subunits of archaeal RNase P (Boomershine et al, 2003; Sidote & Hoffman, 2003; Numata et al, 2004; Takagi et al, 2004; Boomershine, 2005; Kakuta et al, 2005; Wilson et al, 2005) and the availability of a reliable reconstitution assay will now expedite efforts to
illustrate the parallels and differences in the mechanisms of protein-aided RNA catalysis in RNase P from different domains of life and shed light on their ancestry.
Table 3.1 Characteristics of the protein subunits of *Pfu* RNase P

<table>
<thead>
<tr>
<th>Name</th>
<th>Gene ID</th>
<th>pI</th>
<th>Predicted mass, Da</th>
<th>Observed mass*, Da</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pop5</td>
<td>PF1378</td>
<td>10.0</td>
<td>13,708</td>
<td>13,708</td>
</tr>
<tr>
<td>Rpp21</td>
<td>PF1613</td>
<td>10.3</td>
<td>14,172</td>
<td>14,170</td>
</tr>
<tr>
<td>Rpp29</td>
<td>PF1816</td>
<td>10.2</td>
<td>14,953</td>
<td>14,954</td>
</tr>
<tr>
<td>Rpp30</td>
<td>PF1914</td>
<td>9.7</td>
<td>24,363</td>
<td>24,365</td>
</tr>
</tbody>
</table>

* The actual molecular masses of all four proteins were measured by electrospray ionization mass spectrometry. With the exception of Rpp29, the N-terminal Met is removed from the three other *Pfu* Rpps during overexpression in *E. coli*. Therefore, the predicted masses are calculated for the N-terminal Met-lacking variants in the case of Pop5, Rpp21 and Rpp30.
Table 3.2 Oligonucleotide primers used for cloning the RNA and protein subunits of *Pfu* RNase P

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PfuRPR-F</td>
<td>5'-TAGGCGAGGGGGCTG-3'</td>
</tr>
<tr>
<td>PfuRPR-R</td>
<td>5'-GC GAATTCCTAGGCGAGGGGGCTATAG-3'</td>
</tr>
<tr>
<td>PfuPOP5-F</td>
<td>5'-AGTGAGCGTGCAAAACCTTACCTCCTAC-3'</td>
</tr>
<tr>
<td>PfuPOP5-R</td>
<td>5'-GC GAATTCACCCGAATTGAGACAG-3'</td>
</tr>
<tr>
<td>PfuRPP21-F</td>
<td>5'-GCTAAATAACATGAGAAAAAGAAAAAAAGAGAATG-3'</td>
</tr>
<tr>
<td>PfuRPP21-R</td>
<td>5'-GC GAATTCATATTTCCATTTTTTTTTTTCTCTC-3'</td>
</tr>
<tr>
<td>PfuRPP29-F</td>
<td>5'-TGGCGTAACAGCGAAGAAGTGAGAATAG-3'</td>
</tr>
<tr>
<td>PfuRPP29-R</td>
<td>5'-GC GAATTCATTACCGCAACCTCCTCAGACGTCTC-3'</td>
</tr>
<tr>
<td>PfuRPP30-F</td>
<td>5'-GCTGGTGGTGCATATGGTTGGAAGTTTGTAGAGATG-3'</td>
</tr>
<tr>
<td>PfuRPP30-R</td>
<td>5'-GC GAATTCATTATAAGACGTCTCAGAATACCAGTGATATAA AAG-3'</td>
</tr>
</tbody>
</table>

**Note:** The italicized sequences indicate the restriction sites introduced for cloning and linearization of template (for *in vitro* transcription). The nucleotide in bold font (in PfuRPP30-R) refers to a silent substitution that eliminates an internal *Eco*RI site. The underlined sequences indicate regions of complementarity, a requisite for overlap PCR using these oligonucleotides (see text for details).
Figure 3.1 SDS-PAGE analysis illustrating the overexpression in *E. coli* and the homogeneity of the purified preparations of the four protein subunits of *Pfu* RNase P. UI and I refer to total *E. coli* crude lysates prepared from uninduced and induced cultures, respectively. F refers to the final purified preparation. M indicates molecular weight standards.
Figure 3.2 Functional reconstitution of *Pfu* RNase P with the RNA and protein subunits either alone or in various combinations. On the left panel, assays were performed in 30 mM Mg$^{2+}$ for 2 min at 55 °C with 500 nM *E. coli* ptRNA$^{\text{Tyr}}$ as substrate using holoenzymes reconstituted with 50 nM of *Pfu* RPR and 125 nM of the individual protein subunits (or C5 protein). M refers to ptRNA$^{\text{Tyr}}$ processing markers generated with *in vitro* reconstituted *E. coli* RNase P. On the right panel, all possible combinations of protein subunits in *Pfu* RNase P were assayed in 40 mM Mg$^{2+}$ for 3.5 h at 55 °C with 500 nM ptRNA$^{\text{Tyr}}$ using holoenzymes reconstituted with 50 nM of *Pfu* RPR and 500 nM of the individual protein subunits.
Figure 3.3 Activities of partially and fully assembled \textit{Pfu} RNase P. The complete \textit{Pfu} RNase P holoenzyme can be built using two core RNP complexes as the minimal initiating units: \textit{Pfu} RPR+Pop5+Rpp30 (A,C) and \textit{Pfu} RPR+Rpp21+Rpp29 (B,D). Magnesium (A, B) and protein:RNA ratio (C, D) titrations. Relative activities (E) and turnover numbers (F) of partially and fully assembled \textit{Pfu} RNase P complexes under optimal (magnesium and protein:RNA ratio) conditions. Mean values of initial velocities were calculated using data from at least two independent experiments. Error bars reflect mean deviation.
A. Relative activity (%) of [Mg^{2+}] mM

B. Relative activity (%) of [Mg^{2+}] mM

C. Relative activity (%) of Protein:RNA

D. Relative activity (%) of Protein:RNA

E. Relative activity (%) of Protein combinations tested with Pfu RPR

F. Protein combinations tested with Pfu RPR | $k_{cat}$ (min$^{-1}$) | Relative activity (%)
--- | --- | ---
Pop5+Rpp30 | 0.43 ± 0.03 | 4.4
Pop5+Rpp30+Rpp29 | 0.37 ± 0.01 | 3.7
Pop5+Rpp30+Rpp21 | 0.38 ± 0.01 | 3.9
Pop5+Rpp29 | 0.035 ± 0.007 | 0.4
Rpp21+Rpp29 | 0.041 ± 0.003 | 0.4
Rpp21+Rpp29+Rpp30 | 0.105 ± 0.002 | 1.1
Rpp21+Rpp29+Pop5 | 9.87 ± 0.29 | 100
Figure 3.4 Scheme of hierarchical assembly of *Pfu* RNase P.
CHAPTER 4

DISCUSSION

4.1 Summary

In the two described studies above, we have provided information on the role of proteins which aid RNA catalysis in both bacterial and archaeal RNase P. In *E. coli* (bacterial) RNase P, we have employed a site-specific footprinting strategy to position bacterial Rpp on the tertiary structure model of *E. coli* RPR and thereby establish a model of the bacterial RNase P holoenzyme. Furthermore, we have extended this model to also depict an enzyme-substrate (ES) complex of bacterial RNase P (RPR+Rpp bound to ptRNA) by exploiting the extensive literature on specific interactions between ptRNA and bacterial RNase P (both RPR alone and the holoenzyme). In *Pfu* (archaeal) RNase P, we have established a robust reconstitution assay and also obtained a possible hierarchical assembly map for it. Furthermore, steady-state kinetic studies have provided evidence for (i) cooperative binding of *Pfu* Rpps to their cognate RPR, and (ii) the ability of *Pfu* Rpps to lower the Mg$^{2+}$ requirement for *Pfu* RPR catalysis.
4.2 Molecular modeling of the three-dimensional structure of the bacterial RNase P holoenzyme

By converting *E. coli* Rpp into a site-specific nuclease, we mapped the potential contact sites between *E. coli* Rpp and its cognate RPR. This footprinting study revealed that (i) *E. coli* Rpp is located at the catalytic domain of *E. coli* RPR, and (ii) both the highly conserved RNR motif and the unique left-handed crossover loop in bacterial Rpp are proximal to most of the conserved nucleotides in bacterial RPRs (Tsai et al, 2003). The model, which was constructed with our data, suggests that bacterial Rpp plays a direct role in RNase P catalysis by directly contacting the catalytic center of bacterial RPR and that this active site is generated by interactions between conserved residues in the two subunits of bacterial RNase P.

Furthermore, we have also built a three-dimensional model of the enzyme-substrate (ES) complex for bacterial RNase P using three specific contact sites between ptRNA substrates and either bacterial RNase P RNA or protein subunit (Tsai et al, 2003). First, the two Cs in the CCA sequence at the 3′ end of ptRNA base pair with G292 and G293 (in L15) of bacterial RPR (Kirsebom & Svärd, 1994; Oh & Pace, 1994; Svärd et al, 1996). Second, the T stem-loop of ptRNA directly contacts specific nucleotides in P9 and P11 of *E. coli* RPR (Loria et al, 1998). Lastly, crosslinking studies position the leader sequence of ptRNA in the cleft region of bacterial Rpp (Niranjanakumari et al, 1998; Fig. 2.4). Notably, our model of the bacterial RNase P ES complex is also consistent with another important substrate recognition feature. The universally conserved A248 and A249 (in J5/15) nucleotides in *E. coli* RPR is known to be proximal to the -1 position in ptRNA and play an important role for substrate binding and catalysis of bacterial RNase P (Christian et
al, 1998; Christian and Harris, 1999; Siew et al, 1999, Zahler et al, 2003). In our model,
conserved adenosines in the J5/15 region are proximal to the -1 position of ptRNA
substrate, a feature that was not used as a constraint during molecular modeling.

The crystal structures of both types A and B bacterial RPR were solved recently.
Remarkably, these structures are highly similar to the proposed tertiary structure model of
bacterial RPR from previous studies by Prof. Eric Westhof, our collaborator at
IBMC-CNRS, Strasbourg, FRANCE (Massire et al, 1998). The excellent agreement
between the crystal structure and tertiary structure model of bacterial RPR augurs
favorably for the computer-aided models of both the RNase P holoenzyme and the ES
complex. In fact, two recent studies report moderate refinements of the ES complex while
confirming our model. Buck et al. (2005) used in-gel phosphorothioate-iodine
footprinting assay to identify the contact sites between bacterial RPR and Rpp in both
types A and B bacterial RNase P. Their footprinting data position the Rpps on RPR in a
manner consistent with our model (Fig. 4.1A; Buck et al, 2005). Moreover, Rueda et al.
(2005) have used time-resolved fluorescence resonance energy transfer (tr-FRET) to
study the interaction between the leader sequence of ptRNA and the protein subunit in the
context of bacterial RNase P holoenzyme (Fig. 4.1B; Rueda et al, 2005). Their data have
established that the leader sequence between -5 to -8 is not only being positioned in the
cleft of bacterial Rpp but that it is greatly stretched by bacterial Rpp. This observation
refines our model in that it alters the leader conformation in the cleft of bacterial Rpp. In
our model, the 5′ leader of ptRNA while contacting the bacterial Rpp is in a relaxed
manner; however, in the refined version, the leader sequence is stretched as a straight line
while contacting with bacterial Rpp (Fig. 4.1B). Furthermore, Rueda et al. (2005) also
propose the possibility of the interaction between leader sequence of ptRNA and the highly conserved RNR motif in the bacterial Rpp (Fig. 4.1B). This postulate remains to be tested.

4.3 Functional reconstitution of the RNase P holoenzyme from Pyrococcus furiosus, a thermophilic archaeon

By establishing a robust reconstitution procedure for the Pfu RNase P holoenzyme, we made three major observations. First, we were able to demonstrate multiple turnover by the reconstituted Pfu RNase P holoenzyme, a first for any in vivo reconstituted archaeal or eukaryal RNase P. Second, in the course of identifying the minimal catalytically active complexes for Pfu RNase P, we have established two separate minimal core complexes and proposed a possible hierarchical assembly map for Pfu RNase P. The minimal cores of Pfu RNase P are composed of Pfu RPR and two non-overlapping Pfu Rpp combinations. The protein-protein and RNA-protein interactions in the minimal core complexes are consistent with data from previous yeast two- or three-hybrid data and NMR spectroscopic studies (Boomershine, 2005; Jiang & Altman, 2001; Jiang et al, 2001; Houser-Scott et al, 2002; Hall & Brown, 2004; Kifusa et al, 2005; Wilson et al, 2006). Third, we have ascertained that one role of the protein subunits of Pfu RNase P is to decrease the Mg$^{2+}$ requirement for catalysis by Pfu RPR. This observation is reminiscent of one of the roles of protein subunit in bacterial RNase P. Due to the high homology between the protein subunits of archaeal and eukaryal RNase P, it is conceivable that insights on Pfu RNase P catalysis and the availability of a robust
reconstitution assay will pave the foundation for dissecting role(s) of protein cofactors in eukaryal RNase P as well.

4.4 Conclusions

In one possible phylogenetic tree depicting the evolution of RNase P, the last common universal ancestor might have housed an RNase P which was made up of only an RNA subunit even if protein synthesis had already begun (Fig 4.2). With subsequent changes in the pre-biotic soup, it is conceivable that the ancestral RNase P, bereft of protein, evolved two alternative routes to adapt to the environment. In the bacterial lineage, the RNA catalyst recruited one protein cofactor; by the archaeal/eukaryal lineage bifurcation stage, at least four proteins were drafted. With increasing complexity, eukaryal RNase P might have subsequently recruited up to six more proteins. This hypothesis stems from a comparison of the sequence and structure of both RPR and Rpps. The predicted secondary structures of RPRs from all three domains of life are similar with consensus sequences present in all of them (Frank et al, 2000; Haas et al, 1996; Massire et al, 1998). Extant RPRs might have evolved from a common ancestral RNA molecule. The high sequence homology of Rpps among archaea and eukarya suggests that these Rpps might also share a common origin (Hall & Brown, 2002). However, the absence of sequence homology between bacterial Rpps and archaeal/eukaryal Rpps supports an independent origin for the bacterial Rpp.

The main focus of this dissertation has been the elucidation of the roles of protein cofactors in facilitating RNA catalysis. While the study on bacterial RNase P has shed detailed insights in the unique structural features of the bacterial Rpp that permit its
multiple roles in RNase P catalysis (e.g., improving substrate binding, lowering Mg$^{2+}$ requirement) (Reich et al, 1988; Tallsjo & Kirsebom, 1993; Crary et al, 1998; Niranjanakumari et al, 1998; Kurz & Fierke, 2002; Hsieh et al, 2004; Buck et al, 2005), such details are not yet available for the archaeal Rpps. However, the establishment of a robust reconstitution system for archaeal RNase P has already yielded insights which has led us to infer that independent protein-assisted solutions to enhance RNA catalysis in distinct evolutionary lineages have actually converged on the same themes.
Figure 4.1 Comparison of two different models for the interactions between a bacterial Rpp and the 5' leader sequence of a ptRNA substrate. The 5' leader backbone trajectory colored in gray is from this study (Chapter 2) and in green is from the refined model (Rueda et al, 2005). Different motifs are also colored in the bacterial Rpp, the RNR motif is in blue and the central β-sheet is red. The cleavage site is indicated by a hollow arrow. The hydrophobic amino acids in the central cleft (F16, F20, and Y36, based on B. subtilis numbering), which possibly interact with the leader, are drawn in stick representation. The figure and legend are reproduced and adapted, respectively, from Rueda et al. (2005).
Figure 4.2 Cladogram depicting hypothetical evolution of RNase P

LUCA: Last Universal Common Ancestor
CHAPTER 5

ONGOING STUDIES AND FUTURE DIRECTIONS

5.1 Bacterial RNase P

5.1.1 Two-site binding model to understand the interactions between bacterial RPR and Rpp

Our tertiary structural model of the bacterial RNase P holoenzyme with and without its substrate has enabled us to speculate on the potential roles of various conserved residues in bacterial Rpp to catalysis. We propose two RNA-binding sites in bacterial Rpp: one for interacting with its cognate RPR and the other for binding to a ptRNA substrate. We postulated that the highly conserved RNR motif (in $\alpha_2$) and the central cleft region are responsible of interacting with the cognate RPR and the leader sequence of ptRNA, respectively.

While the sequence homology among all bacterial Rpps is low, several residues are highly conserved. Analysis of the mutant derivatives of those highly conserved residues has shown that most of them result in defective holoenzymes (Gopalan et al, 1997). Furthermore, a genetic complementation analysis reveal that a combination of mutations involving two conserved residues results in complete lose of cell viability (Jovanovic et
al, 2002). These studies established the functional significance and the possible cooperation among conserved residues. However, their exact roles remain unclear.

To gain further insights into the role of conserved residues in bacterial Rpp, we have overexpressed and purified various mutant derivatives in which mutations are located at both $\alpha_1$ and $\alpha_2$ regions. These mutant derivatives were further examined with qualitative methods (e.g., RNase T1/V1 protection assay and crosslinking experiments) to investigate if they have defects either in holoenzyme or ES complex formation.

To more quantitatively assess the effect of mutations in bacterial Rpp on RNase P catalysis, we are pursuing both gel-shift and pre steady-state kinetics approaches to analyze the effect of mutations on (i) the binding constant for the interaction of *E. coli* RPR and Rpp, and (ii) the substrate binding constant to the *E. coli* RNase P holoenzyme. These studies are ongoing and will continue to be the focus of studies on bacterial RNase P in the near future.

### 5.1.2 What is the contribution of the left-handed crossover in bacterial Rpp to RNase P catalysis?

Another interesting observation from the proposed tertiary structure model is the unique left-handed crossover region in *E. coli* Rpp. During construction of the computer-aided three dimensional model, two-thirds of the distance constraints (obtained from our footprinting studies) were consistent with our structure, while the remaining one-third failed to show good agreement (Tsai et al, 2003). The disagreements were in data generated from Fe-EDTA footprinting agents localized to the left-handed crossover region. Since the sequence alignment of the left-handed crossover region shows poor
sequence homology among all the bacterial Rpps, the functional contributions of the left-handed crossover (if any) might stem from its structural alterations both before and after binding to its cognate RNA subunit. To evaluate the importance of this loop, especially the need for flexibility or appropriate orientation of certain residues, it might be worth considering certain structural alterations. First, the loop could be disrupted by introducing a single cleavage site for protease recognition (e.g., genenase) in the loop. Second, the flexibility of the loop could be decreased by either crosslinking via two cysteines introduced in the loop or by shortening the length of the loop. These studies have been initiated.

5.2 Archaeal RNase P

5.2.1 Role of protein subunits in archaeal RNase P catalysis

We have established the experimental conditions for reconstituting a robust Pfu RNase P holoenzyme and have also shown that Pfu Rpps could aid RNase P catalysis by decreasing the Mg^{2+} requirement. An immediate goal is to carefully dissect the mechanistic role of Pfu Rpps in Pfu RPR catalysis. Do these proteins contribute to RNA folding, substrate binding, and chemical cleavage? By exploiting the ability to either fully or partially reconstitute Pfu RNase P, we can delineate the contribution of each (or a set of) Pfu Rpps to Pfu RPR catalysis.

As mentioned in Chapter 3 (Page 85), the affinity of the substrate for native Mth RNase P is at least 1000-fold higher than that observed with Mth RPR alone (Pannucci et al, 1999; Andrews et al, 2001). It is reasonable to expect then that we would see an enhancement of substrate binding affinity when we compare Pfu RPR and Pfu RNase P
holoenzyme. By comparing the $K_m$ values for fully and partially reconstituted $Pfu$ RNase P, we should then be able to deduce which protein combination or individual protein contributes to substrate binding. Detailed pre-steady state and steady state kinetic studies on both the partially and fully reconstituted archaeal RNase P holoenzymes would provide the foundation for understanding the kinetic mechanism for archaeal RNase P catalysis, including an appreciation of the rate-limiting step. Furthermore, the functional contribution of individual $Pfu$ Rpps to various steps such as substrate binding, chemical cleavage or product release should also become evident.

5.2.2 Substrate recognition

The mechanisms of substrate recognition by bacterial and eukaryal RNase P are different as evident from their requirements for a minimal substrate (Frank & Pace 1998; McClain et al 1987; Schon, 1999). The subunit composition and characteristics of archaeal RNase P make it a mosaic of bacterial and eukaryal RNase P: the secondary structures of archaeal RPR resemble bacterial counterparts, while archaeal Rpps are homologous to eukaryal Rpps (Frank et al, 2000; Haas et al, 1996; Hall & Brown, 2002; Massire et al, 1998). How does archaeal RNase P recognize its substrates? With a robust reconstitution in hand for $Pfu$ RNase P, we decided to identify the minimal substrate requirement for $Pfu$ RNase P. We are currently examining if various mutant derivatives of ptRNA$^{Gin}$, in which different structural elements have been altered by substitution or deletion, can serve as substrates for bacterial and archaeal RNase P. Preliminary results from turnover number calculations reveal that archaeal RNase P recognizes the minimal substrate of bacterial RNase P, which is not recognized by eukaryal RNase P (Yuan &
Altman, 1995). Some other results suggest that archaeal RNase P is a hybrid in terms of substrate recognition in that it displays attributes typical of both bacterial and eukaryal RNase P. However, more detailed studies (ongoing) are needed to draw firm conclusions.

### 5.2.3 Molecular modeling of archaeal RNase P holoenzyme and ES complex

To understand structure-function relationships in *Pfu* RNase P at atomic detail, we need a high-resolution structure of the *Pfu* RNase P holoenzyme complex. In the last two years, the structures of all four archaeal Rpps have been determined. Although a long-term goal would be to solve the structure of the archaeal RNase P holoenzyme complex, the use of biochemical approaches (e.g., Fe-EDTA footprinting, chemical/enzymatic probing) could provide insights into RNA-protein interactions in the *Pfu* RNase P holoenzyme (albeit at much lower resolution). This information together with the tertiary structure of bacterial RPR, which is closely related to archaeal RPR, might make it possible to build a computer-aided three dimensional model of the archaeal RNase P holoenzyme akin to our successful undertaking with bacterial RNase P (Chapter 2).
REFERENCES


Beebe JA, Kurz JC, Fierke CA (1996) Magnesium ions are required by *Bacillus subtilis* ribonuclease P RNA for both binding and cleaving precursor tRNA<sup>Asp</sup>. *Biochemistry. 35*, 10493-505.


Boomershine WP (2005) Ph. D. Thesis, The Ohio State University, Columbus, OH.


Christian EL, Kaye NM, Harris ME (2000) Helix P4 is a divalent metal ion binding site in the conserved core of the ribonuclease P ribozyme. RNA 6, 511-519.


Crary SM, Niranjanakumari S, Fierke CA (1998) The protein component of Bacillus subtilis ribonuclease P increases catalytic efficiency by enhancing interactions with the 5' leader sequence of pre-tRNA<sup>Asp</sup>. Biochemistry 37, 9409-9416.


Kufel J, Kirsebom LA (1996) Different cleavage sites are aligned differently in the active site of M1 RNA, the catalytic subunit of *Escherichia coli* RNase P. *Proc. Natl. Acad. Sci. 93*, 6085-6090.


Kurz JC, Niranjanakumari S, Fierke CA (1998) Protein component of *Bacillus subtilis* RNase P specifically enhances the affinity for precursor-tRNA\textsubscript{Asp}. *Biochemistry 37*, 2393-400.

Kurz JC, Fierke CA (2002) The affinity of magnesium binding sites in the *Bacillus subtilis* RNase P x pre-tRNA complex is enhanced by the protein subunit. *Biochemistry 41*, 9545-58.


