STRUCTURE AND INTERACTIONS OF ARCHAEAL RNASE P PROTEINS: RPP29 AND RPP21

DISSEDITION

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

By
Yiren Xu, B.S.
Ohio State Biochemistry Program

The Ohio State University
2010

Dissertation Committee:
Mark P. Foster, Advisor
Venkat Gopalan
Charles E. Bell
Richard P. Swenson
ABSTRACT

RNase P is a ubiquitous and essential ribonucleoprotein complex, responsible for 5’ leader sequence removal during tRNA maturation. It is the RNA component that performs the catalysis, making it a ribozyme. As such, RNase P is considered a remnant of a putative RNA-centric world. During evolution from bacteria to eukarya, RNase P typically exhibits loss of RNA structures, coupled with the reduced activity of the naked RNA, and concomitant increase in protein content. This transition makes RNase P an excellent model system to address how proteins could gradually take over the catalytic activity from the RNA along the evolutionary journey from the RNA- to the present protein centric-world. Bacterial RNase P is best understood after three decades of effort. The structures of both the RNA and the single protein cofactor have been determined by X-ray crystallography and NMR spectroscopy. Thorough biochemical studies have revealed how the bacterial RNA subunit appropriately positions catalytic Mg$^{2+}$ ions to catalyze water-mediated nucleophilic attack on the phosphodiester bond, and how the protein subunit assists the RNA structurally. On the contrary, eukaryotic RNase P has been quite a challenge to work with because of its complexity (one RNA and 9 to 10 proteins) and the lack of a robust reconstitution assay for functional studies. Because of evolutionary divergence, the results from bacterial RNase P studies can not be readily transferred to understanding the eukaryal counterpart. On the other hand, archaeal RNase P is composed of one RNA and 5 protein subunits, each of which has a
eukaryotic homolog. Four of them function in pairs (RPP21-RPP29 and RPP30-POP5). Importantly, the resemblance between archaeal and eukaryotic RNase P regarding their sequences and interaction implies that the knowledge gained from the simpler archaeal RNase P can be reasonably projected onto the eukaryotic enzyme. We used solution NMR spectroscopy to determine the structure of the protein-protein complex comprising *Pyrococcus furiosus* RPP29 and RPP21. We found that the protein-protein interaction is characterized by coupled folding of secondary structural elements that participate in interface formation. In addition to detailing the intermolecular contacts that stabilize this 30-kDa binary complex, the structure identifies a surface rich in conserved basic residues likely vital for recognition of the RNA subunit and/or precursor tRNA. Furthermore, enzymatic footprinting experiments allowed us to localize the RPP21-RPP29 complex to the specificity domain of the RNA subunit. These findings provide valuable new insights into mechanisms of RNP assembly and serve as important steps towards a 3D model of this ancient RNP enzyme.

Prior to the determination of the structure of wild type *Pfu* RPP29-RPP21 complex, we also used NMR spectroscopy to determine the solution structure of a variant RPP29Δ17-RPP21V14 complex. An inadvertent Ala to Val mutation at residue 14 in RPP21 led to severe signal loss in the spectra collected on the complex formed from RPP29 wild type and RPP21V14. To circumvent this problem, we explored different N-terminal deletion derivatives using NMR spectroscopy, ITC and an *in vitro* reconstitution assay. RPP29Δ17 was discovered to be the minimal domain for RPP21 (RPP21V14) binding and was used in structure determination of RPP29Δ17-RPP21V14. This structure of the complex serves as a very good representative of the wild type RPP29-RPP21, revealing the structural details of the binding interface between RPP29 and
RPP21 for the first time. Moreover, the work on this variant saved us much effort during structure determination of wild type *Pfu* RPP29-RPP21 complex.

Finally, binding-coupled protein folding observed in both RPP29 and RPP21 directed our research into thermodynamic characterization of this binary complex using ITC. These experiments revealed a large excess \( \Delta C_p \) in support of the NMR-observed induced fit occurring during complex formation. To quantitatively correlate the \( \Delta C_p \) to the extent of folding upon binding, we investigated the ion and proton linkage effects which can also contribute to \( \Delta C_p \). Interestingly, though two salt bridges have been proposed based on the structure, high salt concentration was not seen to impair the RPP21-RPP29 interaction. Instead, the hydrophobic interactions promoted at high salt concentration resulted in tighter binding. In addition, we found that this ion effect does not contribute significantly to \( \Delta C_p \). ITC experiments performed in the buffers with different ionization enthalpy indicated that two protons are released upon binding. Using an empirically determined method, we find that the corrected \( \Delta C_p \) reports 50 residues that fold during complex formation, consistent with the extent of folding observed in the NMR studies.
献给

我挚爱的

父亲徐公达 和 母亲姚雪琴
I would like to express my gratitude to all the people that have made this work possible. First and foremost I would like to thank my advisor Dr. Mark Foster for trusting me on such a wonderful project that I have been in love with even through frustrations and difficulties. His constant encouragement and support is one of my biggest driving powers.

The Foster lab has been my second family, whose countless help and support is most precious to me. From the first day I entered the lab as a rotational student to present, I very much enjoy and appreciate this science-exciting but also very friendly environment, warm and comfortable as home. I am grateful to all members of the Foster lab, past and present, especially to Dr. Ross Wilson, the mentor of my rotation and a great colleague in the lab, for his leading me into the field of NMR spectroscopy and always being there to support and help me. I am forever indebted to Dr. Craig McElroy and Dr. Carlos Amero, who taught me so much about NMR spectroscopy. I would like to thank Ian Kleckner and Dr. Sri Vidya Oruganti for valuable discussions that gave me new ideas and helped me out of difficult situations.

I felt extremely lucky to have had Dr. Venkat Gopalan and his lab as our collaborators. Dr. Venkat Gopalan's knowledge of RNase P and expertise on biochemistry is invaluable. I would like to express my gratitude to his great help whenever I need it and especially his guidance on the Pfu RNase P project. I really enjoy
the interactions between our lab and the Gopalan lab. My special thanks go to Dr. Hsin-Yue Tsai, Dr. Dileep Pulukkunat, I-Ming Cho and Wen-Yi Chen for their priceless support on the biochemical studies of archaeal RNase P. I also would like to thank Dr. Lien Lai who gave me solid training in my rotation in the Gopalan lab and taught me everything I know about molecular cloning.

I would like to thank Dr. Charles Bell and especially Dr. Richard Swenson for serving on my graduate committee and providing valuable suggestions throughout the course of this work.

I want to acknowledge CCIC staff Drs. Charles Cottrell and Chunhua Yuan for their training of me and invaluable assistance.

Finally, my deepest gratitude goes to my loving family, my parents and my husband Qian Bian (卞迁), whose unconditional love and support always keep me company and keep me going forward.
VITA

2003-Present: Ph.D. Biochemistry. The Ohio State University, Columbus, Ohio.

1999-2003: B.S. Biology. Fudan University, Shanghai, China.

PUBLICATIONS


FIELDS OF STUDY

Major Field: Biochemistry
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td>vi</td>
</tr>
<tr>
<td>VITA</td>
<td>viii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xiii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xv</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xix</td>
</tr>
<tr>
<td>CHAPTER 1</td>
<td></td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>RNase P</td>
<td>1</td>
</tr>
<tr>
<td>Bacterial RNase P</td>
<td>3</td>
</tr>
<tr>
<td>Eukaryotic RNase P</td>
<td>4</td>
</tr>
<tr>
<td>Archaeal RNase P</td>
<td>5</td>
</tr>
<tr>
<td>The RNA world</td>
<td>7</td>
</tr>
<tr>
<td>NMR spectroscopy</td>
<td>8</td>
</tr>
<tr>
<td>Sample preparation</td>
<td>9</td>
</tr>
<tr>
<td>Optimization of conditions for protein NMR spectroscopy</td>
<td>10</td>
</tr>
<tr>
<td>Preparation of a complex</td>
<td>11</td>
</tr>
<tr>
<td>Resonance assignments</td>
<td>13</td>
</tr>
<tr>
<td>Backbone torsion angle restraints</td>
<td>16</td>
</tr>
<tr>
<td>NOE-derived distance restraints</td>
<td>17</td>
</tr>
<tr>
<td>Structure determination</td>
<td>19</td>
</tr>
<tr>
<td>Dissertation overview</td>
<td>20</td>
</tr>
<tr>
<td>CHAPTER 2</td>
<td></td>
</tr>
<tr>
<td>SOLUTION STRUCTURE OF A PROTEIN COMPLEX: COUPLED FOLDING OF RPP21-RPP29 IN Pyrococcus furiosus RNase P</td>
<td>41</td>
</tr>
</tbody>
</table>
CHAPTER 4
CHARACTERIZATION OF Pfu RPP29 N-TERMINAL DELETION MUTANTS: MINIMAL DOMAIN OF Pfu RPP29 FOR ASSEMBLY WITH Pfu RPP21
INTRODUCTION
MATERIAL AND METHODS
Protein preparation
Limited proteolysis identified the folded core of RPP29
NMR spectroscopy
Isothermal titration calorimetry (ITC)
Reconstitution assays
RESULTS AND DISCUSSION
Limited trypsin proteolysis
Pfu RPP29 derivatives
NMR assays
ITC
Reconstitution assays
CONCLUSION
CHAPTER 5
DISSECTING BINDING-COUPLED FOLDING IN THE Pfu RPP21-RPP29 INTERACTION BY ISOTHERMAL TITRATION CALORIMETRY
INTRODUCTION
MATERIAL AND METHODS
Protein preparation
Isothermal titration calorimetry
Measurement of $\Delta C_p$ by ITC
Ion linkage
Proton linkage
Tables of thermodynamic parameters obtained from ITC experiments
RESULTS AND DISCUSSION
Enthalpy-entropy compensation
Temperature-independence of $\Delta C_p$
Ion linkage
Proton linkage
LIST OF TABLES

Table 1.1 3D experiments for backbone assignments.....................................................40
Table 2.1 Structural statistics for the wild type *Pfu* RPP29-RPP21 complex.................75
Table 3.1 Steady-state kinetics and thermodynamics of RPP29WT and RPP29Δ17........108
Table 3.2 Structural statistics for the *Pfu* RPP29Δ17-RPP21V14 complex....................109
Table 4.1 Oligonucleotide primers used for cloning the *Pfu* RPP29ΔX derivatives......126
Table 4.2 Relative activity of partially and fully reconstituted RNase P.......................126
Table A.1 Intermolecular NOEs identified from chemical shifts in the 13C-filtered/edited
    NOESY spectrum recorded on [U-13C,15N]-RPP29(*) and unlabeled RPP21...............154
Table A.2 Intermolecular NOEs identified from chemical shifts in the 13C-filtered/edited
    NOESY spectrum recorded on [U-13C,15N]-RPP21(*) and unlabeled RPP29..............156
Table A.3 Intermolecular NOEs identified by iterative structure-based assignment
    (SANE) in 13C-separated NOESY spectra recorded on [U-13C,15N]-RPP29(*) and
    unlabeled RPP21...........................................................................................................158
Table A.4 Intermolecular NOEs identified by iterative structure-based assignment
    (SANE) in 13C-separated NOESY spectra recorded on [U-13C,15N]-RPP21(*) and
    unlabeled RPP29...........................................................................................................160
Table A.5 Intermolecular NOEs identified from chemical shifts in the 13C-filtered/edited
    NOESY spectrum recorded on [U-13C,15N]-RPP29Δ17(*) and unlabeled RPP21V14...168
Table A.6 Intermolecular NOEs identified from chemical shifts in the 13C-filtered/edited
    NOESY spectrum recorded on [U-13C,15N]-RPP21V14(*) and unlabeled RPP29Δ17...169
Table A.7 Intermolecular NOEs identified by iterative structure-based assignment (SANE) in $^{13}$C-separated NOESY spectra recorded on [U-$^{13}$C,$^{15}$N]-RPP29Δ17(*) and unlabeled RPP21V14.................................................................................................................................170

Table A.8 Intermolecular NOEs identified by iterative structure-based assignment (SANE) in $^{13}$C-separated NOESY spectra recorded on [U-$^{13}$C,$^{15}$N]-RPP21V14(*) and unlabeled RPP29Δ17.................................................................................................................................171

Table E.1 Table of the thermodynamic parameters measured to obtain $\Delta C_p$ in 10 mM KCl ...................................................................................................................................211

Table E.2 Table of the thermodynamic parameters measured at different salt concentrations at 55°C...................................................................................................................................212

Table E.3 Table of the thermodynamic parameters measured at different salt concentrations at 10°C...................................................................................................................................213

Table E.4 Table of the thermodynamic parameters measured to obtain $\Delta C_p$ in 150 mM KCl...................................................................................................................................213

Table E.5 Table of the thermodynamic parameters measured in buffers with different ionization enthalpies at pH 6.7, 55°C...................................................................................................................................214
LIST OF FIGURES

Figure 1.1 Function of RNase P...............................................................................................23
Figure 1.2 The secondary structures of RNase P RNA in all domains of life.........................24
Figure 1.3 The proposed mechanism of RNase P based on biochemical studies...............25
Figure 1.4 Crystal structures of bacterial RNase P RNA.......................................................26
Figure 1.5 The crystal structure of bacterial RNase P protein.............................................27
Figure 1.6 A stereo view of the 3D model of the E. coli RNase P holoenzyme-ptRNA complex..........................................................................................................................28
Figure 1.7 Schematic models of eukaryotic RNase P showing the relative arrangements of the RPR and RPPs...........................................................................................................29
Figure 1.8 In vitro reconstitution of Pfu RNase P activity...................................................30
Figure 1.9 Diagram of the steps in structure determination of a protein-protein complex by NMR spectroscopy..................................................................................................................31
Figure 1.10 1D proton NMR spectrum of U-[15N]-RPP29 in complex with unlabeled RPP21 at 600 MHz.........................................................................................................................32
Figure 1.11 The spectral changes reflect the chemical environment difference between free and bound states..................................................................................................................33
Figure 1.12 Sequential backbone assignments using HNCACB.........................................35
Figure 1.13 Chemical shift index (CSI) of RPP29 in complex with RPP21..........................36
Figure 1.14 Backbone torsion angle restraint (φ and ψ) prediction by TALOS are shown in the Ramachandran plot..................................................................................................................37
Figure 1.15 Representative strips from NOESY spectra collected on a protein-protein complex.................................................................................................................................38
Figure 3.6 Chemical shift perturbations illustrate the binding interfaces on RPP21V14 and RPP29Δ17..........................................................100

Figure 3.7 Representative strips from 4D spectrum collected on the ILV-RPP29Δ17 in complex with the unlabeled RPP21V14..........................................................101

Figure 3.8 Superimpositions of RPP29 and RPP21........................................102

Figure 3.9 Overlay of 15N HSQC spectra of RPP21V14 and RPP21WT in the free state ..................................................................................103

Figure 3.10 Representative strips from 13C-filtered/edited NOE spectra collected on RPP29Δ17-RPP21V14 complex.........................................................104

Figure 3.11 Ensemble of the Pfu RPP29Δ17-RPP21V14 complex..................106

Figure 3.12 Electrostatic potential map of the Pfu RPP29Δ17-RPP21V14 complex...107

Figure 3.13 Structural details of A14RPP21 and I24RPP29 ........................................108

Figure 4.1 Sequence alignment of selected RPP29 homologs from Archaea and Eukarya ..................................................................................................................122

Figure 4.2 Mass spectrometric analysis of trypsin-digested Pfu RPP29........123

Figure 4.3 NMR binding assay of Pfu RPP29 N-terminal deletion derivatives........124

Figure 4.4 In vitro reconstitution assay of the fully reconstituted Pfu RNase P...........125

Figure 5.1 Binding isotherms of RPP29 titrated into RPP21..............................145

Figure 5.2 Enthalpy-entropy compensation in the RPP21-RPP29 interaction........146

Figure 5.3 Temperature dependence of the thermodynamics of the RPP21-RPP29 interaction........................................................................................................147

Figure 5.4 Electrostatic potential map of the Pfu RPP21-RPP29 complex........148

Figure 5.5 Ion linkage in the Pfu RPP21-RPP29 complex.................................149

Figure 5.6 Ion linkage and ΔCp........................................................................150

Figure 5.7 Proton linkage in the Pfu RPP21-RPP29 complex.............................151

Figure 5.8 Hydrogen bonding in the RPP21-RPP29 complex.............................152
Figure 5.9 $T_s$ obtained from a fit of temperature dependent $\Delta G$ to the $\Delta C_p$ incorporated Gibbs equation...............................................................................................................153

Figure A.1 Strips from $^{13}$C-filtered / edited NOE spectra recorded on [U-$^{13}$C,$^{15}$N]RPP29 bound to unlabeled RPP21.......................................................................................................................161

Figure A.2 Strips from $^{13}$C-filtered / edited NOE spectra recorded on [U-$^{13}$C,$^{15}$N]RPP21 bound to unlabeled RPP29.......................................................................................................................165

Figure A.3 Strips from $^{13}$C-filtered / edited NOE spectra recorded on [U-$^{13}$C,$^{15}$N]RPP29 $\Delta 17$ bound to unlabeled RPP21V14.................................................................172

Figure A.4 Strips from $^{13}$C-filtered / edited NOE spectra recorded on [U-$^{13}$C,$^{15}$N]RPP21V14 bound to unlabeled RPP29$\Delta 17$.................................................................173

Figure E.1 Representative isotherms of titrations of RPP29 into RPP21 over a temperature range from 10$^\circ$C to 45$^\circ$C in the standard ITC buffer with 10 mM KCl......215

Figure E.2 Representative isotherms of titrations of RPP21 into RPP29 in the standard ITC buffer with different concentration of KCl at 55$^\circ$C.................................................................217

Figure E.3 Representative isotherms of titrations of RPP21 into RPP29 in the standard ITC buffer with different concentration of KCl at 10$^\circ$C.................................................................218

Figure E.4 Representative isotherms of titrations of RPP21 into RPP29 in the standard ITC buffer with 150 mM KCl at temperatures ranging from 10$^\circ$C to 55$^\circ$C.....................219

Figure E.5 Representative isotherms of titrations of RPP21 into RPP29 in buffers with different ionization enthalpies at pH 6.7, 55$^\circ$C.................................................................220
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Å</td>
<td>Angstrom</td>
</tr>
<tr>
<td>°C</td>
<td>degrees Celsius</td>
</tr>
<tr>
<td>CSI</td>
<td>Chemical Shift Index</td>
</tr>
<tr>
<td>CSP</td>
<td>Chemical Shift Perturbation</td>
</tr>
<tr>
<td>Da</td>
<td>dalton(s)</td>
</tr>
<tr>
<td>δ</td>
<td>chemical shift in ppm</td>
</tr>
<tr>
<td>ΔC&lt;sub&gt;p&lt;/sub&gt;</td>
<td>heat capacity change</td>
</tr>
<tr>
<td>ΔG</td>
<td>free energy change</td>
</tr>
<tr>
<td>ΔH</td>
<td>enthalpy change</td>
</tr>
<tr>
<td>ΔS</td>
<td>entropy change</td>
</tr>
<tr>
<td>FPLC</td>
<td>fast protein liquid chromatography</td>
</tr>
<tr>
<td>g</td>
<td>gram(s)</td>
</tr>
<tr>
<td>h</td>
<td>hour(s)</td>
</tr>
<tr>
<td>HPLC</td>
<td>high performance liquid chromatography</td>
</tr>
<tr>
<td>HSQC</td>
<td>Heteronuclear single quantum correlation spectroscopy</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl-βd-thiogalactopyranoside</td>
</tr>
<tr>
<td>ITC</td>
<td>Isothermal titration calorimetry</td>
</tr>
<tr>
<td>k</td>
<td>kilo</td>
</tr>
<tr>
<td>K</td>
<td>kelvin</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>$K_A$</td>
<td>equilibrium association constant</td>
</tr>
<tr>
<td>$K_D$</td>
<td>equilibrium dissociation constant</td>
</tr>
<tr>
<td>L</td>
<td>liter</td>
</tr>
<tr>
<td>m</td>
<td>milli</td>
</tr>
<tr>
<td>µ</td>
<td>micro</td>
</tr>
<tr>
<td>M</td>
<td>moles per liter</td>
</tr>
<tr>
<td>MHz</td>
<td>megahertz</td>
</tr>
<tr>
<td>min</td>
<td>minute(s)</td>
</tr>
<tr>
<td>$Mja$</td>
<td><em>Methanocaldococcus jannaschii</em></td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>$Mth$</td>
<td><em>Methanothermobacter thermoautotrophicus</em></td>
</tr>
<tr>
<td>n</td>
<td>nano</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>NOE</td>
<td>nuclear Overhauser effect</td>
</tr>
<tr>
<td>NOESY</td>
<td>nuclear Overhauser effect spectroscopy</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</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>$Pho$</td>
<td><em>Pyrococcus horikoshii</em></td>
</tr>
<tr>
<td>ppm</td>
<td>parts per million</td>
</tr>
<tr>
<td>ptRNA</td>
<td>precursor transfer RNA</td>
</tr>
<tr>
<td>RMSD</td>
<td>root mean square deviation</td>
</tr>
<tr>
<td>RNase P</td>
<td>ribonuclease P</td>
</tr>
</tbody>
</table>
RNP    ribonucleoprotein
RPP    RNase P protein
RPR    RNase P RNA
s      second(s)
SDS    sodium dodecyl sulfate
TROSY Transverse relaxation optimized spectroscopy
CHAPTER 1

INTRODUCTION

RNase P

Ribonuclease P (RNase P) is a ubiquitous and essential ribonucleoprotein (RNP) complex that is primarily responsible for removing the 5' leader sequence of precursor tRNA (ptRNA) during tRNA maturation (Fig. 1.1).[1-3] Across all domains of life, RNase P is composed of one RNA subunit (RNase P RNA, or RPR) and a varying number of protein subunits (RNase P Proteins, or RPPs). There is one RPP in bacteria, at least five in archaea, and nine in eukarya (Fig. 1.2). The discovery that RNase P is a ribozyme was first reported by the laboratory of Dr. Sidney Altman in 1983, who showed that the catalytic activity resides in the RNA moiety of RNase P in both *E. coli* and *B. subtilis*.[4] The ribozyme activities have also been demonstrated in archaea[5] and eukaryal RNase P[6] by *in vitro* ptRNA processing assays. The ribozyme activity of the RPR alone can only be achieved under elevated monovalent and divalent ion concentrations,[4-6] suggesting an important involvement of metal ions in catalysis. The cleavage of the ptRNA occurs via an in-line nucleophilic attack on the scissile bond phosphorus by a Mg$^{2+}$-bound H$_2$O molecule,[7] resulting in the mature tRNA with a 5' phosphate and the leader sequence with a 3' hydroxyl (Fig. 1.3). Aside from the catalytic role, metal ions may also stabilize the active tertiary structure of the RPR by helping to overcome
electrostatic repulsion, and thus mediating otherwise unfavorable RNA-RNA interactions. However, the RPPs are absolutely required in vivo[8], and enhance catalysis under near physiological conditions in vitro likely by facilitating RPR folding, substrate recognition, and decrease in the Mg$^{2+}$ requirement.[9-13]

To remove the 5' leader of the diverse pool of ptRNAs present in cells, RNase P specifically interacts with three crucial structural elements of ptRNA: 1) the conserved T stem-loop via the S-domain,[14] which is one of the two independently folded domains in RNase P; 2) the acceptor stem through the C-domain,[15] and 3) the 3'-CCA sequence through base-pairing with G292 and G293 in the L15 region of the RPR.[16] Other RNAs that contain a ptRNA-like 5' end, have also been reported to be substrates of RNase P, including transfer-messenger RNA (tmRNA),[17] bacterial operon RNAs,[18; 19] riboswitches,[20] phage regulatory RNAs,[21] and signal recognition particle RNAs.[22]

Because of its versatility in substrate recognition, RNase P has been exploited as a rare and powerful tool for down-regulation of gene expression in eukaryotes. This is achieved by degrading the targeted mRNAs. To accomplish this, an External Guide Sequence (EGS) is designed to be complementary to the target mRNA, and form a ptRNA-like structure. RNase P (RNA) is tethered to the 3’ end of the EGS. Once EGS interacts with the targeted mRNA, a ptRNA-like shape forms and will be recognized and cleaved by the tethered RNase P, leading to the decrease of the targeted mRNA, and therefore down-regulation of the gene expression of the mRNA. This RNase P-based technology of regulation of gene expression has been reviewed in Reference [23].
Though RNase P is essential in all domains of life, significant difference exists between bacterial and human RNase Ps in term of their compositions and structures of subunits, especially in RPPs. Therefore, RNase P is a target for novel antibiotics. Researches have been focused on developing antibacterial drugs that specifically target the bacterial RNase P without impairing the human enzyme activity.[24]

Bacterial RNase P

The best studied RNase P enzymes are from bacteria, and are composed of one large RPR and one small protein cofactor.[25] Based on the primary sequence and secondary structure of the RPRs, bacterial RNase P can be classified into two distinctive types represented by *Escherichia. coli* (type A) and *Bacillus. subtilis* (type B) (Fig. 1.2). [26] In both types, the RPR has been shown to be composed of two independently folding domains, termed the specificity and catalytic domains (S domain and C domain, respectively), held together by interdomain RNA-RNA contacts.[27] To date, crystallographic structures have been reported for two S domains and two full-length RPRs, from bacterial RNase P of type A and B.[28-31] Five distinct and universally conserved regions (CRI-CRV) constitute the structural core of the RPR (Fig. 1.2); and Type-A and Type-B RPR adopt a very similar 3D shape despite differences in secondary structures (Fig. 1.4). The structures of three homologous bacterial RPPs have been solved by crystallography or NMR spectroscopy, and all adopt an α-β sandwich fold with a topology αββαβα (Fig. 1.5).[32-34] The central cleft, where generally a row of three exposed aromatic residues are located, is involved in interactions with the 5'-leader sequence of the ptRNA.[35] Although a crystal structure of the intact bacterial RNase P holoenzyme has not yet been reported, models of A- and B-type RNPs have been built.
using a wealth of information from biochemical and genetic studies (Fig. 1.6). [9; 36; 37]

**Eukaryotic RNase P**

The least understood RNase P is that from eukarya. It consists of one relatively small RPR in size compared to the bacterial RPR and at least nine RPPs.[38] Notably, none of the RPPs share sequence similarity with the single bacterial RPP. Therefore, insights gained from the bacterial RNase P can not be easily extrapolated to the eukaryotic enzyme. Human RNase P RNA exhibits weak activity in an *in vitro* cleavage assay at low pH, which prevents the degradation of the RPR during long assays.[6] Even under conditions with higher ion strength, the cleavage rates were 5-6 orders of magnitude lower than that of the bacterial RPR, which is most likely due to the loss of important RNA elements for maintaining active tertiary structure of RPR. Compared to the bacterial RPR (350-400 nts typically), the eukaryotic RPR (~300 nts) retains the conserved core, but lacks paired regions P15/P16/P17 and P18 in the C-domain (Fig. 1.2). The predicted secondary structure of the S-domain is also quite different, featuring missing P13/P14 or P10.1 in comparison to type A or type B bacterial RPR. The loss of these RNA elements might prevent the eukaryal RPR from folding into its most active conformation, leading to its incompetence in function. The weak catalysis exhibited by the human RPR *in vitro* is presumably rescued *in vivo* when properly assembled with all the RPPs.

An explanation for the poor activity of the human RPR may be that the critical RNA-RNA interactions that stabilize bacterial RPRs may have been taken over by the RNA-protein interactions in eukaryotic enzymes. Yeast two-hybrid and three-hybrid assays have established an extensive interaction network among all the subunits in yeast and
human RNase Ps (Fig. 1.7),[39-41] indicating 1) reciprocal interactions between RPP21 and RPP29 (in both human and yeast), and between RPP30 and POP5 (only observed in human); 2) RPP21, RPP29, RPP30 and RPP38 directly interact with human RPR; the RPP29-RPR interaction is also confirmed in yeast. Recently, a crystal structure was reported on the complex made from a modified P3 domain and two RPPs, Pop6 and Pop7, of yeast RNase P, suggesting possible functions of such an RNA-protein complex in stabilization of the holoenzyme’s structure and in substrate interactions.[42]

Archaeal RNase P

Phylogenetic and biochemical studies have revealed that archaeal RNase P is a compositional intermediate between the bacterial and eukaryotic counterparts. Archaeal RNase P enzyme can also be categorized into two groups based on their RPR sequences: A and M (Fig. 1.2).[26; 43] Archaeal A-type RPR (e.g., Pyrococcus furiosus, or Pfu) is similar to the bacterial A-type in terms of secondary structures and reported in vitro catalytic activity;[5; 26] in contrast, the archaeal M-type RPR more closely resembles the eukaryotic RPR and has shown no catalytic activity on its own, although it has been shown to cleave a substrate tethered in cis.[44] Though no high-resolution structure of the archaeal RPR is available, secondary-structure similarities and the presence of universally conserved nucleotides in the active site suggest that the archaeal RPR fold (especially the C domain) resembles that of the bacterial RPR. However, the increased RPP content in archaeal RNase P also hints that like the eukaryal enzyme, RNA-protein interactions are likely to have replaced some of the intramolecular RNA-RNA interactions present in the bacterial RPR.[45]
All identified archaeal RPPs share sequence similarity with human RPPs, but not with the single bacterial RPP. This suggests a common ancestor between archaeal and eukaryotic RNase P, at least in terms of the RPPs.[46] Four protein subunits have been identified to associate with archaeal (\textit{Methanothermobacter thermoautotrophicus}, \textit{Mth}) RNase P and named after their human counterparts: RPP21 (PF1613 in \textit{Pfu}), RPP29 (PF1816), RPP30 (PF1914), and POP5 (PF1378). The structures of each have been solved from different archaeal organisms by NMR spectroscopy, X-ray crystallography, or both. These studies on the isolated proteins revealed the structures of the RPPs to fall within common nucleic acid binding protein families: an Sm-like fold (RPP29),[47-50] a zinc ribbon (RPP21),[51; 52] a TIM-barrel (RPP30),[53] and an RRM-like fold (POP5)[54]. A putative ribosomal protein L7Ae is able to increase the optimal reaction temperature from 55°C to 70°C in \textit{Pho} RNase P \textit{in vitro}. Moreover, it shares a significant homology to human RPP38, making it a fifth RPP in archaeal RNase P.[55]

Yeast two-hybrid studies conducted with archaeal RPP21, RPP29, RPP30 and POP5 indicated the presence of two sets of binary complexes: RPP21-RPP29 and RPP30-POP5.[56; 57] Reconstitution assays performed with \textit{Pfu} RNase P have shown that either protein pair is sufficient to activate the RNA enzyme at lower ion concentrations, while no single protein can rescue the RPR under the same conditions (Fig. 1.8).[58] In addition, \textit{Methanocaldococcus jannaschii} RPP30-POP5 complex has been shown to significantly increase the $k_{obs}$ of ptRNA-Mja RPR conjugate self-processing by providing the functional groups to make up the active site in RPR.[44] Although these studies established a role for RPP binary pairs in enhancing the catalytic activity of archaeal RPRs, the mechanistic basis for their action is poorly understood without useful structural models of the archaeal RNase P holoenzyme.
The RNA world

The 1989 Nobel Prize was shared by Thomas Cech and Sidney Altman for the discoveries of RNA-based catalysis in the Tetrahymena group I intron\cite{59} and RNase P\cite{4}, respectively. These findings were revolutionary not only because they were the first discovered catalytic RNAs, but also because they provided the first evidence to support the RNA world hypothesis proposed by Francis Crick and Leslie Orgel in the 1960s (reviewed in Ref. \cite{60}). These findings showed that RNA is not only a passive carrier of genetic information between DNA and protein in the central dogma of molecular biology, but can actively participate in molecular function by performing enzymatic catalysis. Unlike the group I intron and other self-cleaving RNAs discovered since, RNase P is of more fundamental interest since it is a true enzyme, capable of multiple turnover reaction, and it is ubiquitous -- its function is required in all organisms.

However, despite similar function across all domains of life, the composition of RNase P varies from bacteria through eukarya. As a ribozyme, the complexity of the RPR determines the enzymatic activity residing in the naked RNA.\cite{1; 4; 45; 61} To date, bacterial RPR has been shown to be most active, followed by archaeal RPR and eukaryotic RPR. The loss of the RNA elements is also inversely correlated with increased protein content. The single small bacterial RPP only accounts for ~10% of the entire mass of bacterial RNase P, while ~50% and ~70% of the mass of archaeal and eukaryotic RNase P are contributed by their RPPs, respectively. This relationship between the RNA complexity, activity and protein content suggests that RNase P is likely a remnant from the putative RNA world, where catalysis was performed by RNA. Thus, with a few exceptions, nature appears to have compensated for the loss of the RNA
structure by adding more proteins while retaining function along the path of molecular evolution.

Assuming the RNA world hypothesis is correct, it would seem that only a few, but notable catalytic RNAs have survived the crucial evolutionary journey from RNA world to protein world. These include RNase P,[60] the ribosome[62] and splicesome,[60] all of which are essential for life in all organisms. Our research focuses on RNase P to address how protein cofactors might have taken over the structural and functional attributes of RNAs during the evolution from the RNA-centric world to the present protein-centric world.

**NMR spectroscopy**

NMR spectroscopy is one of the two methods, along with X-ray crystallography, for structure determination of large biomolecules at atomic resolution. Compared to X-ray crystallography, NMR spectroscopy has the following advantages. 1) NMR-derived solution structure is generally considered to be a better representation of biomolecules in their native conditions, in which artifacts from crystal packing forces can be avoided. 2) Though both methods require highly concentrated samples in solution to begin with, X-ray crystallography requires ordered crystals that can diffract X-rays to high resolution, which can sometimes be extremely hard to obtain. 3) NMR spectroscopy can be used to study transiently formed complexes with weak affinity, which might not be suitable for crystallization. In addition to yielding structural information at atomic resolution, NMR is a powerful technique for studying the dynamic behavior of molecules on a wide range of timescales from picosecond through minutes to hours. This property greatly helps to understand molecular motions that are important for folding, catalysis and interactions of
biomolecules.

On the other hand, the high quality of NMR spectra are severely limited by the size of macromolecules. Large molecules (> 25 kDa) intrinsically give more signals in the NMR spectra, resulting in spectral overlap. Slower tumbling of large molecules results in a shorter transverse relaxation time (T$_2$), leading to broad linewidths and signal loss before detection. To overcome the T$_2$-related problems, TROSY (TRansverse Relaxation Optimized Spectroscopy)[63] and uniform deuterium labeling[64; 65] have been introduced into NMR spectroscopy; these methods retain only the narrowest component of a coupling multiplet and reduce spin-spin interactions by diluting protons, respectively. Last but not least, data collection at elevated temperature increases the tumbling rate of molecules, thus decreasing T$_2$.[66]

In the following text, we focus on use of NMR for structure determination of a protein-protein complex consisting of a heterodimer formed from two distinct proteins with strong affinity for each other (K$_D$ $\leq$ 10$^{-7}$ M). The overall approach is shown by the scheme in Fig. 1.9. The procedure can be categorized into three parts: 1) sample preparation, 2) NMR spectra collection, and 3) data analysis and structure calculation.

Sample preparation

NMR spectroscopy of proteins relies on the magnetic properties of the nuclei of $^1$H, $^{13}$C and $^{15}$N. Despite the naturally abundant $^1$H isotope, high level of enrichment with $^{13}$C and / or $^{15}$N is required for heteronuclear edited NMR spectra. Following well documented procedures, $^{13}$C and $^{15}$N can be readily incorporated into proteins by recombinatly expressing in E. coli and growing cells in defined minimal media with $^{15}$NH$_4$Cl and $^{13}$C-glucose as the sole nitrogen and carbon sources, respectively. When
studying large proteins and their complexes (MW > 40-kDa), we can uniformly deuterate the protein to overcome the faster $T_2$ relaxation by diluting $^1$Hs with $^2$Hs and thus reducing spin-spin relaxation.[64; 65] However, uniform deuterium labeling also silences the carbon-attached protons. Thus, site specific protonation on a perdeuterated protein is more practical for structure determination. This strategy has been shown effective with ILV (Ile, Leu, Val) methyl-protonated, $^2$H, $^{13}$C, $^{15}$N-labeled protein.[67; 68]

All the aforementioned labeling methods can be applied to both component of a protein heterodimer. However, forming a complex with one protein labeled and the other protein unlabeled can help to further simplify NMR spectra and data analysis. Heteronuclear-edited NMR spectra collected on such a complex will only record the signals from the labeled protein, in the context of the complex, while the signals from its unlabeled partner will not be observed. Therefore, this approach for structure determination of a protein-protein complex requires a minimum of two samples, with each protein labeled and bound to its unlabeled partner.

**Optimization of conditions for protein NMR spectroscopy**

Before collecting a series of NMR spectra, it is important to identify an optimal experimental condition that ensures a stable sample and sufficient signal-to-noise (S/N) ratio in NMR spectra. Parameters such as buffer, pH, salt, other additives and temperature should be taken into consideration. For a protein complex, the condition should also guarantee and/or promote the formation of a stable complex. A large number of different NMR buffers can be quickly screened by the hanging drop technique, in which a small amount of protein(s) is mixed with the target buffer, and checked for precipitation over time (1-2 days).[69] NMR spectra are also a powerful diagnostic tool
to evaluate buffer conditions, because NMR signals are very sensitive to the local chemical environment and molecular weight. For instance, the linewidth in an NMR spectrum is a great reporter on the size of molecules, indicating aggregation propensity of the sample. Widely dispersed proton signals indicate a well folded protein (Fig. 1.10), whereas clustered signals in the spectrum reflects an unfolded protein since all protons experience a similar environment. This NMR screening can usually be carried out with unlabeled protein. Because of the small amount of sample required for the hanging drop method, many conditions can be examined simultaneously, and some unfriendly conditions, like those that cause sample precipitation, can be excluded immediately. After preliminary screens, more detailed assessment can be achieved from NMR spectra, including folding, aggregation and feasibility of the future NMR studies. Though recording 1D $^1$H NMR spectra is quick (15 minutes per sample), it requires much more material limiting throughput. Therefore, condition optimization is generally started with the hanging drop method, covering a wide range of conditions, followed by NMR spectra for fine-tuning.

**Preparation of a complex**

Macromolecular NMR methods are primarily $^1$H-detected. Thus, they only record the proton spectrum. However, because of the way the magnetization is manipulated during the experiment, the proton signals are modulated by the heteronuclear nuclei ($^{15}$N or $^{13}$C) that are attached to the protons. This allows the recorded proton signal to be correlated to the signals of the coupled nuclei. By adding the second dimension ($^{15}$N frequency), the 1D spectrum can be resolved according to the chemical shifts ($\omega$) of the attached $^{15}$N. A 2D $^{15}$N-edited HSQC spectrum records chemical shifts of each amide proton and its attached nitrogen (as present in the backbone and some side chains). Therefore, 2D
HSQC spectra provide a conventional fingerprint for each residue (except proline) in the protein.

Upon complex formation, the chemical environment of the residues in the interface changes, and the NMR signals of these residues also change accordingly. As an example, when unlabeled RPP21 is titrated into U(uniformly)-[^15N]-RPP29, signals from free RPP29 disappear, with concomitant appearance of a new set of resonances corresponding to RPP21-bound RPP29. The chemical shift change between free and bound states is defined as $\Delta \omega$. No further changes are seen in the HSQC spectrum of U[^15N]-RPP29 after reaching a molar ratio of 1:1 of RPP29 and RPP21 (Fig. 1.11). These observations indicate a slow exchange between free and bound protein is slow (the exchange rate $k \ll \Delta \omega$), and a tight complex formed between PP29 and RPP21 with stoichiometry of 1:1. Alternatively, a weak complex ($K_D > 20 \mu M$ or higher) will typically give rise to signals that show intermediate ($k \approx \Delta \omega$) to fast ($k >> \Delta \omega$) exchange.[70] In these cases, instead of observing disappearance of free signals and appearance of bound signals (two signals when half bound), in the fast exchange regime, signals of free and bound protein will average, and appear as weighted average signals between free and bound (i.e., one set of signals when half bound at the midway between free and bound chemical shifts). As the NMR chemical shift time scale is defined by the difference between the frequency of nuclei in the two states, two states exchanging at intermediate timescale ($k \approx \Delta \omega$) leads to broad or vanishing signals, and should be avoided if possible by optimizing sample conditions.

Aside from informing on the affinity of the interaction and on stoichiometry, comparison of 2D HSQC spectra between free and bound states of a protein also identifies the likely binding interface. Residues in or close to the binding interface
experience the largest chemical environment change, thus exhibit the largest chemical shift perturbations (CSP) between free and bound signals. It is noteworthy that over-analysis of CSPs may lead to an overestimate of the extent of the interface because it is impossible to distinguish CSPs that arise from direct interaction from those attributed to local conformational rearrangement upon binding. Nevertheless, CSPs help to highlight the likely binding site, even without the knowledge of the structure, and can be used as restraints in subsequent structure determination.[71]

**Resonance assignments**

To map the CSPs onto the structure of a protein, each peak in a 2D $^1\text{H}-^{15}\text{N}$ spectrum must be assigned to amide $^1\text{H}-^{15}\text{N}$ pair in both free and bound 2D spectra. Sequence-specific assignments in proteins are most commonly obtained by recording a protein of “triple resonance” NMR spectra that allow through-bond correlation of backbone resonances (i.e., $^1\text{H}$, $^{15}\text{N}$, $^{13}\text{C}$). These experiments take advantage of large 1- and 2-bond scalar ($^1J$, $^2J$) $J$ coupling between backbone nuclei for magnetization transfer. Like a 2D spectrum is obtained from many 1D spectra, a third frequency dimension ($^{13}\text{C}$) is added to a 2D $^{15}\text{N}$-edited $^1\text{H}$ spectrum, and the chemical shifts of $^{13}\text{C}$ nuclei that are attached to $^1\text{H}-^{15}\text{N}$ will be recorded to further resolve the spectrum. Table 1.1 lists a set of triple resonance experiments commonly used for backbone assignments; these record chemical shifts of $^1\text{HN}$, $^{15}\text{N}$ and $^{13}\text{C}$. Minimally, backbone assignments might be achieved from high quality 3D HNCO and HNCACB spectra. The HNCO records chemical shifts of amide $^1\text{H}-^{15}\text{N}$ and its proceeding C'. Because $^{15}\text{N}$-edited HSQC spectra also contain signals from side chain $^1\text{H}-^{15}\text{N}$ pairs, HNCO is commonly used to distinguish the backbone signals from side chain peaks in 2D HSQC spectrum. The HNCACB
correlates the preceding \((i-1)\) and intra-residue \((i)\) \(C^\alpha\) and \(C^\beta\) chemical shifts to each backbone amide \(^1\text{H}-^{15}\text{N}\). So in this spectrum, each \(\text{HN}\) shift in an HNCACB spectrum will be correlated to four \(^{13}\text{C}\) frequencies (unless one or both of the residues is a Gly). Due to the relatively low sensitivity of the HNCACB, additional 3D experiments are often required, including HNCA, CBCA(CO)NH and others. In a CBCA(CO)NH spectrum, each amide proton is correlated to its attached \(^{15}\text{N}\) \((i)\) and the \(^{13}\text{C}^\alpha\) and \(^{13}\text{C}^\beta\) shifts of the preceding residue only \((i-1)\), which facilitates distinguishing the inter-residue \(\text{Ca}\) and \(\text{C}\beta\) from intra-residue ones. The HNCA is similar to HNCACB, but only correlates \(C^\alpha\) chemical shifts to \(^1\text{H}-^{15}\text{N}\). The information obtained from CBCA(CO)NH and HNCA are redundant with that from the HNCACB. However, they may provide additional information if the S/N is low in the HNCACB.

In a typical case, sequential backbone assignments are made using 2D strips extracted from a 3D spectrum based on the position of signals in a 2D HSQC spectrum (Fig. 1.12). Using the HNCACB as an example, each strip contains four peaks attached to the same \(^1\text{H}-^{15}\text{N}\) pair. Because of the magnetization transfer pathway in the HNCACB spectrum, the peaks corresponding to \(C^\beta\) have opposite phase to the peaks from \(C^\alpha\). Generally, since the average values of \(^1J_{\text{CaN}} > ^2J_{\text{CaN}}\) and \(^2J_{\text{C\betaN}} > ^3J_{\text{C\betaN}}\), the intra-residue correlations are more intense than the inter-residue correlations. Moreover, with help from the CBCA(CO)NH spectrum which only records the \(^{13}\text{C}\) chemical shifts of the preceding residues, the intra- and inter-residue \(C^\alpha\) and \(C^\beta\) can be clearly distinguished. Thus, in an HNCACB strip, the chemical shifts of inter-residue \(C^\alpha\) and \(C^\beta\) should match the intra-residue \(C^\alpha\) and \(C^\beta\) shifts in a strip from the preceding residue, while its intra-residue \(C^\alpha\) and \(C^\beta\) chemical shifts should match inter-residue \(C^\alpha\) and \(C^\beta\) chemical shifts in another strip from the next residue. This connectivity continues until no predecessor or successor can be found.
Sequence specific assignments is facilitated by the fact that some amino acids have unusual chemical shifts for their C$^\alpha$ and C$^\beta$ resonances. For example, Gly only has C$^\alpha$ at ~45 ppm; Ala, Ser and Thr have unusual C$^\beta$ chemical shifts (Ala C$^\beta$ ~20 ppm, Ser C$^\beta$ ~60 ppm, and Thr C$^\beta$ ~70 ppm). When a certain combination of amino acids is found in the fragment, such as GXT where X means any amino acids, it can be used to locate the fragment in the primary sequence (Fig. 1.12).

Once the backbone assignments are made, side chain assignments can be obtained from additional triple resonance spectra, such as the 3D H(CCO)NH-TOCSY and C(CO)NH-TOCSY. Given their relatively low sensitivity, analogous information can be obtained from 3D HB(CB)HA(CA(CO)NH, HCCH-TOCSY and HCCH-COSY experiments.[72] Importantly, assignment of all possible resonance will greatly assist in obtaining NOE restraints for structure determination.

The process of obtaining resonance assignment also yields valuable structural information. The chemical shifts of backbone atoms are strongly dependent on the character and nature of protein secondary structure. For example, compared to a random coil resonance, C$^\alpha$ experience an upfield shift in the α helical conformation and a downfield shift in the β conformation. As such, the Chemical Shift Index (CSI) compares the backbone chemical shifts of assigned residues to the chemical shifts for a residue in a random coil conformation to determine the α or β propensity of these residues (Fig. 1.16).[73] The predicted secondary structures are generally very reliable, and can be used as a guidance in restraints assignments (See next section).

When making resonance assignments for a complex, the assignments for the free protein can serve as a useful guide. However, unless only very small CSPs are observed, or the CSPs can be followed during titration (typically for a weakly bound
complex with faster exchange), the backbone assignments must be performed on both labeled free proteins and both complexes with one protein labeled and the other unlabeled.

**Backbone torsion angle restraints**

In a similar manner to predicting secondary structure from chemical shifts, this information can be used at a higher level of contribution to obtain restraints on the backbone $\phi$ and $\psi$ angles. One program for doing this is TALOS,[74] which suggests $\phi$ and $\psi$ restraints based on comparison of chemical shifts of H\textalpha, N, C', C$^\alpha$ and C$^\beta$ to those of residues in known structures.[73] TALOS uses a database comprised by a large number of well-determined high resolution structures with known resonance assignments. Typically, TALOS examines $\phi$ and $\psi$ of each residue (N) in a context of 3-residue peptide (N-1, N, N+1), looks for 10 peptides in the database with the highest similarity of chemical shifts and composition (Fig. 1.14). Based on the distribution of these $\phi$ and $\psi$ angles of residues with similar shifts, users can determine whether the chemical shift data are statistically predictive. Alternatively, $\phi$ can be obtained experimentally from measurement of 3-bond coupling $J_{\text{HNH\textalpha}}$, measured in HNHA and use of the Karplus relationship.[75]

**NOE-derived distance restraints**

The majority of restraints employed in structure calculation by NMR spectroscopy are from NOE-derived inter-proton distances. The NOE (Nuclear Overhauser Effect) arises from through-space magnetization transfer between protons within ~5 Å. In general, for macromolecular NMR, 3D NOESY spectra are recorded in which two of the
dimensions correspond to the protons involved in the NOE, and the third is of a $^{13}\text{C}$ or $^{15}\text{N}$ nucleus that is directly bonded to one of the protons. Such experiments are called $^{13}\text{C}$- or $^{15}\text{N}$-edited or -separated NOESY. The assignment of an NOE is based on the chemical shifts of the protons and heteronucleus; therefore, complete assignment of backbone and side chain resonances is important. Because the intensity of NOE peak is inversely proportional to the sixth power of the distance ($r$) between two $^1\text{H}$s (NOE $\sim 1/r^6$),[72] interproton distances can be determined from the magnitude of the NOE, and used as restraints in structure calculation.

In large macromolecules with highly overlapped $^1\text{H}$ signals, unambiguous assignment of a few NOEs can be achieved based on unique chemical shifts, such as those from methyl and aromatic protons. In addition, distinct patterns of NOEs can also be found in secondary structures; for instance, strong NOEs between amide protons of three residues apart in α-helices, and of two residues in adjacent β-sheets. However, in typical macromolecules, it is hard to unambiguously assign most of the NOEs due to the similar chemical shifts of many protons. This process is even more complicated in NOESY spectra collected on a complex, since they contain both intra- and intermolecular NOEs.

To expedite the identification of NOEs, it is common to use structural information to resolve unambiguities in NOE assignments arising from overlap. For this work, we used the semi-automated program SANE (structure-assisted NOE evaluation).[76] As its name indicates, SANE requires a structure (or an ensemble of structures) to serve as a distance filter. The structure can be derived from partial assignments of NMR data and secondary structure information, or from a homology model. In the absence of a tertiary structure model, SANE can also look for certain patterns of NOEs based on secondary
structure information. Typically, SANE first identifies all possible assignments for an NOE based on chemical shifts alone, then removes the unlikely assignments by passing all candidates through a distance filter based on the structures. The reported assignments depend on user-defined distance cutoff and a maximal number of possibilities allowed. For example, four ambiguous assignments will be reported only if users allow four or more possibilities. In my experience, though SANE is not reliable to assign NOEs de novo when no structural information is available, it is a very quick and trustworthy method to assign a large number of NOEs using structure-based distance filters once a reliable structural template is in hand.

An important and challenging requirement in structure determination of a complex is to assign and distinguish a relatively small number of inter-molecular NOEs from the many intra-molecular NOEs. The inter-molecular NOEs can be selectively recorded in heteronuclear filtered/edited NOESY spectra.[77; 78] The filtering step removes coherences of the protons attached to $^{15}$N or $^{13}$C, allowing magnetization from protons attached to other heteronucleus to pass through (Fig. 1.15). The editing step selects for proton coherences attached to $^{15}$N or $^{13}$C. Thus, the experiment only records NOEs from protons not attached to $^{15}$N/$^{13}$C to $^{15}$N- or $^{13}$C-attached protons. When a complex is formed with one protein labeled and the other unlabeled, the recorded NOEs are exclusively intermolecular.

To complement the short range structural information from NOEs and three-bond coupling, two important NMR methods can provide medium range and global information. When a molecule is partially aligned relative to the magnetic field, the anisotropic tumbling of the molecule results in incomplete averaging of dipole-dipole interactions. This residual Dipolar Coupling (rDC) can be measured for each $^1$H-$^{15}$N pair
(bond vector) by comparing \(^1\)H-\(^{15}\)N correlated spectra of macromolecules in isotropic solution and partially in magnetically aligned media.[79; 80]. The measured rDC depends on the average orientations of the H-N bond vector, and thus provides global structural information. Paramagnetic Relaxation Enhancement (PRE) experiments take advantage of paramagnetic groups with unpaired electrons which cause rapid relaxation of neighboring nuclei. By attaching a paramagnetic spin label to a specific site in a protein, the spectral change (intensity and position of the signals) of other residues is a reporter of the spatial relationship between other residues with respect to this residue[81; 82]

**Structure determination**

For structure determination by NMR spectroscopy, we used the Xplor-NIH suite, which performs structure calculations using minimization protocols based on molecular dynamics and stimulated annealing[83] In the absence of a good starting structure template, an extended structure is generated from the primary sequence (Fig. 1.16a). When performing structure calculation on a protein-protein complex, it is often convenient to use a template consisting of a one continuous polypeptide with discrete residue numbers. For example, protein A is numbered from 1 and protein B starts from 501.

Typically, a set of initial structures can be folded from an extended template using unambiguous NOE assignments and backbone torsion angle restraints (Fig. 1.16). Ten to twenty structures with lowest energy or least number of violations are generally used for further evaluation in iterative structure-assisted NOE assignment. Based on this initial ensemble, additional distance restraints derived from NOEs can be assigned manually
or by semi-automated methods, and are included for next round of structure calculation. It is generally a good idea for the calculation to start from the extended template each time so as to prevent bias towards the an intermediate structure, or from the folded structure to encourage production of more accurate conformations, and increase the calculation efficiency. This iterative process can be performed in a semi-automated fashion to expedite the process. Iterative calculations continue until the convergence can not be further improved. A final ensemble of structures is reported, and the quality of the ensemble is evaluated using tools such as them in Xplor-NIH and PROCHECK-NMR. [84] We found it more efficient to perform structure calculations on the individual proteins first. Thus, SANE only needs to assign intramolecular NOEs in 3D NOESY spectra, and Xplor-NIH calculation time for a single protein is also shortened compared to that for the protein-protein complex. Structure calculation of the complex can also start with an extended template, using previously assigned intramolecular NOEs, which allow good convergence of each protein, in combination of unambiguous intermolecular NOEs identified from filtered/edited NOESY spectra. Additional intermolecular NOEs can be assigned by SANE to be included into the next round of calculation once the initial structures of the complex are available. When no further convergence can be achieved, the refinement process is complete.

Dissertation overview

A functional protein-protein interaction between RPP21 and RPP29 has been established for both archaeal[56; 57]and eukaryotic[39; 41] RNase P using yeast two-hybrid assays. The in vitro reconstitution assay has also demonstrated the functional importance of this protein binary complex.[58] The main work described in this
dissertation is the structure determination of the *Pfu* RPP21-RPP29 complex by NMR spectroscopy. During the research, an Ala to Val mutation in *Pfu* RPP21 (termed RPP21V14) initially hindered NMR data collection, and led us to explore the minimal domain of *Pfu* RPP29 required for assembly with RPP21V14 to allow further structure studies (Chapter 4). We used NMR spectroscopy, ITC and *in vitro* reconstitution assays as diagnostic tools to investigate the function of different N-terminal deletion constructs of RPP29, and found that RPP29Δ17 was a suitable construct. We determined the structure of the *Pfu* RPP29Δ17-RPP21V14 complex, and for the first time shed light on the binding interface between two proteins as well as the potential protein-RNA interface (Chapter 3). After identifying this complicating RPP21 A14V mutation, we repeated the structure determination process on a complex formed by the wild type and full length proteins, *Pfu* RPP29-RPP21. Notably, we discovered that binding-coupled folding occurs in both proteins (Chapter 2). To explain the results more clearly, these three chapters are presented in reverse chronological order.

The observation of induced fit upon binding between RPP21 and RPP29 also led us to characterize the *Pfu* RPP29-RPP21 complex from a thermodynamic perspective (Chapter 5). In particular, since a large negative ΔC_p (heat capacity change) is a thermodynamic signature of binding-coupled folding,[85] we performed ITC experiments over a wide range of temperatures (10°C to 45°C) to measure ΔC_p (-1.136 kcal mol⁻¹ K⁻¹) from the temperature dependence of binding enthalpy. To correct the measured ΔC_p for effects of ion and proton linkage, we investigated the thermodynamics of RPP21-RPP29 interaction at different salt concentrations and buffers with different ionization enthalpies. We found that high salt concentrations favors binding, consistent with burial of nonpolar surface, and that ion linkage does not contribute significantly to the measured ΔC_p. We
also found the transfer of two protons from proteins to buffer upon interaction at pH 6.7. 

The corrected $\Delta C_p$ reports that there are 49.4 residues folded when forming the complex, consistent with the 50 residues observed to fold upon binding in NMR studies.

The determination of the structure of *Pfu* RPP29-RPP21 complex is an important step towards understanding the architecture and function of archaeal and eukaryal RNase P, moving us a critical step closer towards a model of the archaeal RNase P holoenzyme.
Figure 1.1. Function of RNase P. The primary role of RNase P is to cleave the 5' leader sequence of precursor tRNA (left), and generate mature tRNA (right).
Figure 1.2. The secondary structures of RNase P RNA in all domains of life. All RPRs have a folded core (a), composed of 5 conserved regions (CRI-CRV). *E. coli* (b) and *B. subtilis* (c) were first demonstrated as ribozymes for bacterial A- and B-type RPR. Archaeal RNase P can also be classified into Type A (d) and M (e), represented by *Pfu* and *Mja*. Eukaryotic RNase P RNA (f) is exemplified by human enzyme. The different numbers of RNase P proteins (RPPs) that are associated with the RPR are indicated in each domain of life.
Figure 1.3. The proposed mechanism of RNase P based on biochemical studies. The in-line nucleophilic attack is performed by a Mg$^{2+}$-coordinated H$_2$O molecule (red). Two other Mg$^{2+}$ ions are also involved in binding of pro-R$_p$ oxygen of the scissile phosphate (adapted from Ref[25]).
Figure 1.4. Crystal structures of bacterial RNase P RNA. The structures of bacterial Type A and B RPR have been solved from *Thermotoga. maritima* (a) and *B. stearothermophilus* (b) RNase P, respectively. Paired regions (labeled in PX, X is the helix number) in the secondary structures (top panel) are colored according to the arrangement of the coaxially stacked helical domains in the crystal structures (bottom panel). The gray regions are disordered, and either not observed (a) or modeled based on the crystal structure of S-domain from *B. subtilis* (b). The figure is adapted from Ref. [25]
Figure 1.5. The crystal structure of bacterial RNase P protein. The structure was determined on *B. subtilis* RPP, which adopts an αββαββαβα fold. A-helices and β strands are labeled sequentially by letters (A-C) and numbers (1-4). Three potential RNA binding interfaces are proposed, 1) a left handed helix containing RNR consensus sequence (red in panel A), 2) a central cleft at the base of which there are a row of three exposed aromatic residues (panel B), and 3) a carboxylate-rich loop (green in panel A) where two Zn$^{2+}$ ions bind. The figure is adapted from Ref. [34]
Figure 1.6. A stereo view of the 3D model of the *E. coli* RNase P holoenzyme-ptRNA complex. Elements in the secondary structure (i.e., P1, paired region 1) are labeled in different colors. The ptRNA and C5 protein (bacterial RPP) are shown in silver and cyan, respectively. The green spheres indicate the positions of 1) the base-pairing between the 3'-ACCA terminus of ptRNA and G292/G293 in P15 of M1 RNA, and 2) the proximity between nucleotide -4 in the leader sequence of ptRNA and residue 50 in C5 protein. The figure is adapted from Ref.[36].
Figure 1.7. Schematic models of eukaryotic RNase P showing the relative arrangements of the RPR and RPPs. The RPPs are shown as ovals and RPRs are shown either in secondary structure diagram (a) or as line (b). In yeast RNase P (a), Pop4 (RPP29) and Pop1p have been shown to interact with the RPR, with Pop1p specifically to P3 of RPR. Where on RRP Pop4 interacts is not clear. In human RNase P (b), P21, P29 and P38 are in contact with H1 RNA (red), while the other proteins are not seen for H1 RNA interaction. The results were from thorough yeast two- and three-hybrid experiments. The figures are adapted from Ref. [41] and [40].
Figure 1.8. *In vitro* reconstitution of *Pfu* RNase P activity. The activity was assayed by adding a known concentration of *E. coli* ptRNA\(^{\text{Tyr}}\), a trace amount of which was internally labeled with [α-\(^{32}\)P]GTP. The reaction was performed at 55°C, and quenched with urea-phenol dye. The substrate (ptRNA\(^{\text{Tyr}}\)) and products (mtRNA\(^{\text{Tyr}}\) and 5' leader) were separated by denaturing PAGE gel electrophoresis based on the size. Reactions performed in *Pfu* RPR alone (1), *Pfu* RPR+C5 protein (2) and *Pfu* RPPs without RPR (3) showed no cleavage. Both fully reconstituted *Pfu* RNase P (4, 19) and *E. coli* RNase P (M) showed strong activity, and indicated the position of the substrate and the products on the gel. The cleavages were also assayed with the different combination of recombinant *Pfu* RPR and RPPs. No single protein with the RPR can boost the activity (5-8). Only two specific protein pairs can rescue the activity, which are RPP30-POP5 (11) and RPP21-RPP29 (12), while the other combination of two RPPs did not show activity (9, 10, 13, 14). The addition of the third protein into either specific protein pair did not significantly increase the activity (15-18), until all four RPPs are present (19). These results confirm the functional importance as protein-protein complexes, RPP21-RPP29, and RPP30-POP5. (adapted from Ref.[58]).
Figure 1.9. Diagram of the steps involved in structure determination of a protein-protein complex (i.e., RPP29-RPP21) by NMR spectroscopy. There are three major steps, sample preparation (black arrows), NMR data collection (red arrows), and data analysis and structure calculation (blue arrows).
Figure 1.10. 1D proton NMR spectrum of U-[\textsuperscript{15}N]-RPP29 in complex with unlabeled RPP21 at 600 MHz. The spectrum of a well folded protein sample is characterized by widely dispersed H signals, according to their distinctive chemistry and local environment. DSS is often used as a reference in the aqueous solution. WATERGATE scheme is generally applied to suppress an otherwise large H\textsubscript{2}O signal.[72] Note that the 1D spectrum contains the protons from both RPP29 and RPP21.
Figure 1.11. The spectral changes reflect the chemical environment difference between free and bound states. In the case of a tight protein-protein complex (slow exchange rate, $k << \Delta \omega$), the 2D HSQC spectrum collected on a half bound U-[\textsuperscript{15}N]-protein contains two sets of signals (b, blue), one of which shares the same chemical shifts with the free state signals (a-c, black). The other set of signals in the half bound protein spectrum is located at new positions, which correspond to the bound state signals (b-c, red). Note that the intensities of the blue peaks are half of the intensities of the black or red peaks, because only half of the proteins are either free or bound. The representative spectra are close-up regions of 2D HSQC spectra collected on free (a), half bound (b) and completely bound (c) U-[\textsuperscript{15}N]-RPP29 in complex with unlabeled RPP21.
Figure 1.12. Sequential backbone assignments using HNCACB. Representative strips are from the HNCACB, which are pulled out based on the attached amide moiety (a). The strips are aligned with $^1$H chemical shifts at x-axis and $^{13}$C chemical shifts at y-axis. Connectivity are made based on the matching chemical shifts of C$^\alpha$ (black) and C$^\beta$ (red) resonances correlated to individual amide resonance. Unique chemical shifts in the linked fragment help to determine the location in the primary sequence, such as GXT (X is any amino acid).
Figure 1.13. Chemical shift index (CSI) of RPP29 in complex with RPP21. Secondary structures (shown as cartoons at the bottom) are predicted based on the chemical shifts of backbone atoms in comparison of their values in random coil conformation.
Figure 1.14. Backbone torsion angle restraint (φ and ψ) prediction by TALOS are shown in the Ramachandran plot.[86] φ and ψ torsion angles of ten peptides, which share the highest similarities of chemical shifts and composition with the target residue (in the context of 3-residue peptide). The cross and circle indicate the mean and standard deviation of predicted φ and ψ value. Nine out of 10 are clustered in the α configuration region, suggesting the φ and ψ of this particular residue also 90% likely adopts an α configuration.
Figure 1.15. Representative strips from NOESY spectra collected on a protein-protein complex. Comparison of strips from $^{13}$C-edited NOESY (yellow) and $^{13}$C-filtered / edited NOESY (cyan) shows that $^{13}$C-edited NOESY spectrum contains both intra- and intermolecular NOEs, while $^{13}$C-filtered/edited NOESY only records the intermolecular NOEs. The red signals in $^{13}$C-filtered/edited NOESY are from incomplete suppression of self- and intramolecular NOEs.
Figure 1.16. Structure convergence during structure calculation. Xplor-NIH starts with an extended template (a) of RPP21, and fold it into 3D structure using NOE-derived distance restraints and backbone torsion angle restraints. The convergence gradually improves when more restraints are included, from 2.4 Å (b) to 0.8 Å (c).
Table 1.1. 3D experiments for backbone assignments. (The table is adapted from Ref. [72])

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Correlations observed</th>
<th>Magnetization transfer</th>
<th>$J$ couplings $^b$</th>
</tr>
</thead>
</table>
| HNCA           | $^1\text{H}^N - ^{15}\text{N}_i - ^{13}\text{C}_\gamma$  
                | $^1\text{H}^N - ^{15}\text{N}_i - ^{13}\text{C}_{\gamma-1}$  | $^1\text{J}_{\text{NH}}$  
                | $^1\text{J}_{\text{NCO}}$  
                | $^2\text{J}_{\text{NC}}$  |
| HN(CO)CA       | $^1\text{H}^N - ^{15}\text{N}_i - ^{13}\text{C}_\gamma$  | $^1\text{J}_{\text{NH}}$  
                | $^1\text{J}_{\text{NCO}}$  
                | $^1\text{J}_{\text{C}_\gamma\text{CO}}$  |
| H(CA)NH        | $^1\text{H}^p - ^{15}\text{N}_i - ^{1}\text{H}^N$  
                | $^1\text{J}_{\text{NH}}$  
                | $^1\text{J}_{\text{NC}}$  
                | $^2\text{J}_{\text{NC}}$  
                | $^1\text{J}_{\text{SH}}$  |
| HNCO           | $^1\text{H}^N - ^{15}\text{N}_i - ^{13}\text{CO}_{\gamma-1}$  | $^1\text{J}_{\text{NH}}$  
                | $^1\text{J}_{\text{NCO}}$  |
| HN(CA)CO       | $^1\text{H}^N - ^{15}\text{N}_i - ^{13}\text{CO}$  
                | $^1\text{J}_{\text{NH}}$  
                | $^1\text{J}_{\text{NC}}$  
                | $^2\text{J}_{\text{NC}}$  
                | $^1\text{J}_{\text{C}_\gamma\text{CO}}$  |
| CBCA(CO)NH     | $^{13}\text{C}_\beta^{,}/^{13}\text{C}_\alpha - ^{15}\text{N}_{i+1} - ^{1}\text{H}_{i+1}$  | $^1\text{J}_{\text{CH}}$  
                | $^1\text{J}_{\text{C}_\beta\text{CO}}$  
                | $^1\text{J}_{\text{C}_\gamma\text{CO}}$  
                | $^1\text{J}_{\text{CO}}$  
                | $^1\text{J}_{\text{SH}}$  |
| CBCANH         | $^{13}\text{C}_\beta^{,}/^{13}\text{C}_\alpha - ^{15}\text{N}_i - ^{1}\text{H}^N$  | $^1\text{J}_{\text{CH}}$  
                | $^1\text{J}_{\text{C}_\beta\text{CO}}$  
                | $^1\text{J}_{\text{C}_\gamma\text{CO}}$  
                | $^1\text{J}_{\text{CO}}$  
                | $^1\text{J}_{\text{NH}}$  |
| HNCACB         | $^{13}\text{C}_\beta^{,}/^{13}\text{C}_\alpha - ^{15}\text{N}_i - ^{1}\text{H}^N$  
                | $^{13}\text{C}_\beta^{,}/^{13}\text{C}_\alpha - ^{15}\text{N}_{i-1} - ^{1}\text{H}_{i-1}$  | $^1\text{J}_{\text{CH}}$  
                | $^1\text{J}_{\text{C}_\beta\text{CO}}$  
                | $^1\text{J}_{\text{C}_\gamma\text{CO}}$  
                | $^1\text{J}_{\text{CO}}$  
                | $^1\text{J}_{\text{NH}}$  |

$^b \ ^1\text{J}_{\text{NH}} \sim 91$ Hz, $^1\text{J}_{\text{NCO}} \sim 7-11$ Hz, $^2\text{J}_{\text{NCO}} \sim 4-9$ Hz, $^1\text{J}_{\text{NCO}} \sim 15$ Hz, $^1\text{J}_{\text{C}_\gamma\text{CO}} \sim 55$ Hz, $^1\text{J}_{\text{CH}}$($^1\text{J}_{\text{C}_\beta\text{CO}}$), $^1\text{J}_{\text{C}_\beta\text{CO}}$) $\sim 140$ Hz, $^1\text{J}_{\text{C}_\alpha\text{C}_\beta} \sim 35$ Hz.
CHAPTER 2

SOLUTION STRUCTURE OF A BINARY PROTEIN COMPLEX:
COUPLED FOLDING OF RPP21-RPP29 IN Pyrococcus furiosus RNase P

INTRODUCTION

The homology between archaeal RPPs and eukaryotic RPPs suggests that archaeal and eukaryotic RNase P have the same ancestor.[87] However, the simpler architecture of archaeal RNase P makes it easier to study the protein-protein and protein-RNA interactions in the holoenzyme, while the fruitful results from the study of the archaeal RNase P will also shed a light on the eukaryotic enzyme. We have focused our effort on the RNase P from the hyperthermophilic archaeon Pyrococcus furiosus (Pfu), taking advantage of the following traits: 1) robust in vitro reconstitution assays for Pfu RNase P have been developed in the Gopalan Lab, which have facilitated functional studies;[88] 2) compared to their eukaryotic counterparts, Pfu RPPs are smaller and more thermostable, making them more amenable to structural study by NMR spectroscopy.

The immunoprecipitation experiments have shown that there are four RPPs associated with the archaeal RPR, termed according to the human homologues: RPP21, RPP29, RPP30 and POP5.[87] All four RPPs belong to fold families that are commonly seen in RNA/DNA binding proteins (Fig. 2.1). The structure of RPP29, the first
structurally characterized RPP, has been determined in *Methanobacterium thermoautotrophicum* (Mth), *Archaeoglobus fulgidus* (Afu) and *Pyrococcus horikoshii* (Pho) by both NMR spectroscopy and X-ray crystallography.[48-50; 89] Only the Sm-like core of RPP29 is structured in solution, while the terminal helices are also stabilized in the crystal structure, probably due to the crystal packing. A difference in the solution and crystal structures was also observed in RPP21, where the N-terminus was observed in a helical conformation in the crystal, but was unstructured in solution while the overall L-shaped fold remains the same.[51; 52] The structure of free POP5 has only been solved for *Pfu* by X-ray crystallography.[54] Interestingly, POP5 shares a very similar RRM topology with the bacterial RPP, though there is no significant sequence similarity between them. The POP5 protein partner RPP30 contains a TIM barrel, in which ten helices and seven β-sheets form a solenoid, curving around to close on itself in a doughnut shape.[53] Recently, the genome search using yeast and human RPPs in *Pho* has found that PH1496 shares 48% and 23% identical residues with the *H. marismortui* (Hma) ribosomal protein L7Ae and human RPP38 (hence, denoted as RPP38). The crystal structure of *Pho* RPP38 shows that it has an α-β sandwich structure, in a strong resemblance to the *Hma* L7Ae.[55] However, whether RPP38 is authentic RNase P protein subunit still remains unclear, though addition of it in reconstitution assays for *Pfu* and *Pho* increase the optimal temperature from 55°C to 70°C.[personal communication with the Gopalan lab][55]

As suggested by the yeast two-hybrid and reconstitution results (Fig. 2.1), two specific protein pairs, RPP21-RPP29 and RPP30-POP5, may assemble into the binary complexes prior to binding to RNA.[56-58] The crystal structure of such a protein-protein complex was first reported for the *Pho* RPP30-POP5 complex (Fig. 2.3).[90] The
interface between RPP30 and POP5 occupies a small surface area and is dominated by hydrogen bonds. Two copies of each proteins are present in the asymmetric unit, forming a heterotetramer. However, the functional importance of the dimerization of the complex is still unclear.

We have determined the NMR-derived solution structure of the 30-kDa complex between RPP29 and RPP21 of Pfu RNase P. This study complements the recently reported crystal structure of the Pho RPP21-RPP29 complex, reveals dynamic features of the proteins and binding-coupled protein folding events, and identifies additional features important for protein-protein and protein-RNA interactions. Furthermore, footprinting studies allow us to map the RPP21-RPP29 complex onto the S domain of the RPR and the RPP30-POP5 complex onto the C-domain of the RPR. The high-resolution structure determinations of the binary complexes are a necessity towards building a 3D model of archaeal RNase P holoenzyme.

MATERIALS AND METHODS

The Pfu RPP29 wild type and RPP21 wild type are referred to as “Pfu RPP29” or “Pfu RPP29WT” and “Pfu RPP21” or “Pfu RPP21WT”, respectively. The N-terminal deletion variants of Pfu RPP29 are referred to explicitly as “Pfu RPP29ΔX”, where ΔX indicates the first X residues are removed by mutagenesis. Pfu RPP21 A14V mutant is denoted as “Pfu RPP21V14”, indicating a Valine at residue 14 instead of an Alanine.

Protein expression and purification

Pfu RPP29
The *Pfu* RPP29/pET-33b plasmid[58] was transformed into *Escherichia coli* BL21 (DE3) Rosetta cells (Novagen). The Rosetta cells supply tRNAs for rare codons in order to enhance the recombinant protein expression. The cells were grown in 2-L flasks in a shaker-incubator at 37°C in Luria-Bertani (LB) or minimal M9 media containing 1 g/L of \(^{15}\)NH\(_4\)Cl and 2 g/L of \(^{13}\)C-glucose as the sole nitrogen and carbon sources, supplemented with 30 \(\mu\)g/L of kanamycin (for pET-33b) and 34 \(\mu\)g/L of chloramphenicol (tRNA helper plasmid in Rosetta cells). Production of the recombinant proteins was induced by addition of 0.5 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) at an \(A_{600}\) of 0.6 and harvested after 4 hours by centrifugation (10 min, 6000xg, room temperature). The cell pellet from 1 L of culture was resuspended in 30 mL of Buffer R [25 mM Tris-HCl, pH 7, 1 mM EDTA (ethylenediaminetetraacetic acid), 0.1 mM PMSF (phenylmethanesulphonylfluoride), and 50 mM DTT (dithiothreitol)], lysed on ice by sonication, centrifuged (15 min, 8000xg, room temperature). The resulting pellet was resuspended in 30 mL of Buffer R containing 7 M urea. The solution was sonicated on ice again, and the cell debris was removed by centrifugation (15 min, 8000xg, room temperature). The supernatant was filtered (0.45 \(\mu\)m) and loaded onto a 5-mL HiTrap SP column pre-equilibrated with buffer R. *Pfu* RPP29 was eluted using a 50-mL linear 10-25% gradient of 25 mM to 2 M KCl under denaturing conditions (25 mM Tris-HCl, pH 7, and 7 M urea), refolded by dialyzing into NMR Buffer [10 mM Tris-HCl, pH 6.7, 10 mM KCl, 0.3 mM ZnCl\(_2\), and 0.02% (w/v) NaN\(_3\)], and concentrated by ultrafiltration with a 5-kDa molecular mass cutoff membrane (Ultra-4, Amicon). Protein concentrations were determined based on the predicted extinction coefficient (37,470 M\(^{-1}\)cm\(^{-1}\), ProtParam) under denaturing conditions (0.02 M Phosphate, pH 7.5, 6 M guanidinium HCl).
We initially used an RPP21 Ala14Val variant (denoted RPP21V14) in our structural studies due to inadvertent selection of a clone with this spurious mutation (See Chapter 3). In order to revert to wild type \textit{Pfu} RPP21, we used “Quick Change” method to amplify the whole plasmid using PCR from the \textit{Pfu} RPP21V14/pET-33b plasmid as the template and gene-specific DNA primer (forward: 5'-GAA AAA AAG AGA ATT \textbf{GCA} AAA GAG AGA ATT \textbf{GAT} ATC TTG TTT AGC-3'; reverse: 5'-GCT AAA CAA GAT \textbf{ATC} AAT TCT CTC TTT TGC AAT TCT TTT TTC-3', where the bold NTs indicate the Val to Ala mutation, and underlined NTs are removal of the \textit{Cal} restriction site for screening prior to automated DNA sequencing. The removal of the \textit{Cal} site also resulted in a silence mutation at Ile18, indicated by italics.) The PCR products were subjected to \textit{DpnI} digestion for 1 hour at 37°C, followed by transformation into \textit{E. coli} DH5α cells. The plasmid encoding \textit{Pfu} RPP21 wild type was purified by Miniprep Kit (Qiagen) and the sequence was confirmed by automated DNA sequencing.

The \textit{Pfu} Rpp21-pET33b plasmid was transformed into \textit{E. coli} BL21 (DE3) Rosetta cells and grown at 37°C in LB or minimal M9 media with 1 g/L of $^{15}$NH$_4$Cl and 2 g/L of $^{13}$C-glucose as the sole nitrogen and carbon sources, supplemented with 30 μg / L of kanamycin and 34 μg / L of chloramphenicol. When the OD$_{600}$ reached 0.6, cells were induced by adding 1 mM IPTG and 50 μM ZnCl$_2$, and the cells were harvested after 4 h by centrifugation (10 min, 6000xg, room temperature). Cells were resuspended in 15 mL of Lysis buffer (50 mM Tris-HCl, pH 7.4, 50 mM KCl, 5 mM imidazole, 6 M guanidine-HCl) and lysed on ice by sonication. Cell debris was removed by centrifugation (1 h, 16000xg, room temperature). The filtered supernatant was loaded onto a 5 mL HiTrap Chelating column (GE Biosciences) charged with 100 mM NiSO$_4$. The His-tagged \textit{Pfu}
RPP21 was eluted by a linear 0-100% gradient of 50-500 mM imidazole under denaturing conditions (50 mM Tris-HCl, pH 7.4, 50 mM KCl, 8 M urea). After refolding in NMR buffer, the His-tagged RPP21 was incubated with thrombin (10 units/mg of Pfu RPP21) at 37°C overnight for removal of the His-tag, resulting in a 122-residue protein consisting of 116 natively encoded residues and 6-residue tail (GGLVPR) at the C-terminus. Then, the protein was purified by C4 reversed-phase HPLC chromatography (Vydac 214TP1010, 1 x 25 cm; 90 ml linear gradient of 25-60% acetonitrile, 0.1% TFA), lyophilized, resuspended in denaturing buffer (50 mM Tris-HCl, pH 6.7, 50 mM KCl, 50 mM DTT, 6 M guanidine-HCl, 0.5 mM ZnCl$_2$) and heated to 70°C for 20 min. Finally, Pfu RPP21 was refolded via dialysis into NMR Buffer and concentrated by ultracentrifugation. Protein concentration was estimated based on the predicted extinction coefficient (16180 M$^{-1}$cm$^{-1}$, ProtParam) under denaturing condition (0.02 M Phosphate, pH 7.5, 6 M guanidinium HCl).

**NMR spectroscopy and resonance assignments**

The Pfu RPP21-RPP29 complex was formed by combining each U-$^{15}$N,$^{13}$C-labeled protein (~1 mM) with an unlabeled protein partner at a 1:1.2 molar ratio to promote full saturation of the labeled protein.[51] To improve the spectral quality, we took advantage of the thermostability of Pfu proteins, and recorded NMR spectra at 55°C so as to promote the molecule tumbling rate, which will result in a slower T$_2$ and a narrower linewidth. The experiments were collected on 600- and 800-MHz Bruker Avance DRX spectrometers (Billerica, MA) equipped with triple-resonance pulse-field gradient probes. NMR spectra were processed and analyzed by NMRPipe,[91] NMRView,[92] and CARA.[93]
Assignments of $^1$H, $^{15}$N and $^{13}$C resonances for each labeled protein in complex with its unlabeled partner were obtained by using data from the following experiments at 600 MHz: $^{15}$N-$^1$H HSQC (heteronuclear single quantum coherence), $^{13}$C-$^1$H HSQC, HNCO, HNCA(CB), CBCA(CO)NH, C(CO)NH-TOCSY (total correlated spectroscopy) ($\tau_m=12$ ms) and H(C)(CO)NH-TOCSY ($\tau_m=12$ ms).[72] Distance restraints were obtained from 3D $^{15}$N-separated NOESY-HSQC ($\tau_m=100$ ms) and $^{13}$C-separated NOESY-HSQC ($\tau_m=100$ ms) spectra recorded at 800 MHz on samples dissolved in 10% and 99.8% D$_2$O, respectively. Intermolecular distance restraints were assigned initially from $^{13}$C-filtered/edited NOESY-HSQC spectra ($\tau_m=100$ ms, 800 MHz, 99.8% D$_2$O) with the two $^1J_{CH}$ filter elements tuned to 120 and 160 Hz.[77; 78]

A heteronuclear {${^1}{H}$-}$^{15}$N NOE experiment was recorded at 600 MHz on [U-$^{15}$N,$^{13}$C] RPP29 bound to unlabeled RPP21 in an interleaved manner with (NOE) and without (no NOE) $^1$H saturation.[94] Heteronuclear NOE values were determined from the ratios of the peak intensities between the two spectra.

**Structure determination**

Distance restraints were derived from the NOE peak intensities, which were calibrated by assigning the median intensity to an interproton distance, $r$, of 2.7 Å and scaling the remaining intensities by $1/r^6$.[92] Restraints for methyl groups and geminal protons were adjusted by adding 0.5 Å to the upper bound. Backbone torsion-angle restraints were obtained from analysis of the backbone chemical shifts via TALOS.[86] Hydrogen bonds were identified based on the secondary structural information and characteristic NOE patterns.
Structure calculations were performed using simulated annealing protocols within the Xplor-NIH software suite.[83] Preliminary analyses of the $^{15}$N,$^{13}$C-edited and $^{13}$C-filtered/edited NOE spectra identified a total of 323 unique (nonredundant) unambiguous NOEs, including 105 intramolecular NOEs in RPP21, 124 intramolecular NOEs in RPP29, and 284 intermolecular NOEs in the binding interface. An initial set of structures of each individual protein was generated using the unique intramolecular NOEs, in addition to short-range NOEs and dihedral angle restraints. These initial ensembles (RPP21 and RPP29, respectively) were then used as templates for iterative computer-aided structure-based NOE assignment (SANE, or structure-assisted NOE evaluation) [76] of the isolated proteins, resulting in assignment of additional intramolecular NOEs. Then, using the NOE and dihedral angle restraints obtained for the refined structure ensembles of the individual proteins, and the 284 intermolecular NOEs, and initial ensemble of the protein-protein complex was generated from an extended template. This initial docked ensemble of the complex served as the template for iterative SANE, yielding an additional 191 intermolecular NOEs. One hundred trials of restraints-driven refinement resulted in an ensemble of ~70 structures with similar restraint energies, from which a final set of 10 structures was selected for analysis and evaluation with Xplor-NIH and PROCHECK-NMR.[84] Surface burial was calculated using the program STC.[95]

**Binding interfaces identified by Chemical Shift Perturbation (CSP)**

The free and bound HSQC spectra of each U-$^{15}$N-labeled protein were recorded at 55°C and 600 MHz, and examined for the chemical shift differences of amide resonances between the free and the bound spectra. Weighted average CSPs were calculated from:
where $f = 8$ for Gly, and $f = 6$ otherwise.\[71; 96\]

**Accession numbers**

Coordinates and restraints have been deposited in the Protein Data Bank (PDB) with accession number 2KI7. NMR resonance assignments have been deposited in the Biological Magnetic Resonance Data Bank (BMRB) with accession number 16266.

**RESULTS AND DISCUSSION**

**NMR spectroscopy of the RPP29-RPP21 complex**

At 25°C, NMR spectrum of free Pfu RPP29 was of poor quality (Fig. 2.4), with broad lines and few peaks, probably due to non-specific aggregation. Similarly, spectrum of Pfu RPP21 recorded at 25°C has broad lines and highly variable peak intensities, suggesting that the protein was poorly folded and/or aggregated at room temperature. Spectra recorded at 55°C yielded generally narrow lines and uniform peak intensities for both proteins, and thus, data were recorded at 55°C for both the free proteins and the protein-protein complexes. The $^{15}$N-edited HSQC spectrum of free Pfu RPP29 contains only 62 of 124 expected backbone resonances (Fig. 2.5), suggesting that only its Sm-like core is folded in the absence of its partner RPP21, consistent with previous studies on RPP29 from Afu[49] and Mth.[47] In free RPP21, 78 backbone amides could be assigned in the free protein, revealing a structured core comprising residues 19 to 105, out of a total of 123 (Fig. 2.5).[51] As unlabeled RPP21 is titrated into $^{15}$N-labeled
RPP29, signals from free RPP29 disappear, with concomitant appearance of a new set of resonances corresponding to RPP21-bound RPP29, until the samples reach a molar ratio of 1:1 of RPP29 and RPP21, beyond which no further change in the spectrum is observed. This behavior corresponds to the slow exchange regime for the equilibrium between the free and bound states and is indicative of tight binding and a 1:1 stoichiometry.

Strikingly, 52 new RPP29 backbone amide resonances are observed in \(^{15}\text{N}\)-edited spectra of the RPP21-bound RPP29. This number corresponds overall to 42% of the RPP29 primary sequence and accounts for nearly all of the resonances that were not observed in spectra of the free protein. Among these 52 new signals, 41 arise from residues at the N-terminus and 8 from the C-terminus, implying that the termini, which are disordered in free state, fold upon binding to RPP21 and play an important role in forming the RPP21 binding interface (Fig 2.6).

Spectra of bound RPP21 also revealed signals not observed in the free protein. Upon binding to unlabeled RPP29, nine new backbone amide resonances could be assigned to residues from the N-terminal helical bundle of the protein. In addition, the amides that exhibited the largest CSPs upon binding RPP29 are primarily from residues clustered at the N-terminal helical bundle of RPP21, including D19, I20, L21, L24, A25, R27, V28, S32, R38, L42 and V46 (Fig. 2.6). These data highlight the RPP29 binding region of RPP21[51] and indicate that the interaction with RPP29 stabilizes the secondary structure in this region of RPP21.

CSPs allowed preliminary identification of the binding interface between RPP29 and RPP21. Detailed characterization of the binding interface was achieved by recording and assigning intermolecular nuclear Overhauser effects (NOEs), which were obtained from
3D $^{13}$C-filtered / edited NOE spectroscopy (NOESY) spectra recorded in 99.8% D$_2$O.[77; 78] Cross peaks in these spectra arise from NOEs between protons not attached to $^{13}$C nuclei (the filter step) to protons attached to $^{15}$C nuclei (the editing step), thereby providing exclusively intermolecular NOEs when one of the proteins is uniformly $^{13}$C-labeled and the other is unlabeled. We recorded $^{13}$C-filtered /edited NOESY spectra on both [U-$^{13}$C,$^{15}$N]RPP29 in complex with unlabeled RPP21 and on [U-$^{13}$C,$^{15}$N]RPP21 in complex with unlabeled RPP29 (Fig. 2.7). These spectra yielded a total of 284 intermolecular NOEs (153 and 131 NOEs from each spectrum).

**Solution structure of the Pfu RPP29-RPP21 complex**

The solution structure of the *Pfu* RPP29-RPP21 complex (Fig. 2.8) was determined by iterative torsion-angle refinement using distance restraints derived from inter- and intramolecular NOEs, hydrogen-bond restraints inferred from secondary-structure information, and torsion-angle restraints from analysis of chemical shifts. The ensemble is well defined for the assigned residues (17-123 of RPP29 and 9-54, 57-81, and 86-104 of RPP21) with a mean root-square deviation (RMSD) of 0.58 Å and 0.87 Å for backbone and heavy atoms, respectively. The stereochemical quality of the structures was high, with 98.1% of the residues adopting φ and ψ angles falling in the most favored and the additionally favored regions of the Ramachandran plot (Table 2.1). The extreme N- and C-termini of both proteins (residues 1-16 and 124-127 of RPP29; residues 1-8 and 105-123 of RPP21) are flexible in the complex, as confirmed by ($^1$H-$^{15}$N heteronuclear NOE data (Fig. 2.9), indicating that these regions are not involved in the interaction with the protein partner.
In the RPP29-RPP21 complex, additional structural elements from each protein could be resolved, which were disordered in their free states. For RPP29, in addition to the signals from the β-barrel core, binding to RPP21 allowed definition of residues from three helices (helix α1, residues 19-23; helix α2, residues 27-31; and helix α3, residues 40-44), an extended strand connecting helix α2 and helix α3 at the N-terminus, and a C-terminal helix (helix α4, residues 117-122). In the context of the RPP21 complex, the packing of several N-terminal residues of RPP29 within the Sm-like core of the protein (I24, T27, R31, H34, V38, K40, and L44) is stabilized, although these intramolecular interactions were insufficient to stabilize the interactions in the free protein. In contrast to the extensive intramolecular contacts made by the binding-stabilized N-terminal residues, the RPP29 C-terminal helix observed in the complex appears to be entirely stabilized by intermolecular contacts to RPP21, as few intramolecular NOEs are observed from this region of the protein to other parts of RPP29. The Sm-like core of RPP29 is essentially unchanged by RPP21 binding, as the RMSD between the free and RPP21-bound structures is within the precision of the ensemble (~1.1 Å for backbone atoms). Thus, it is evident that binding of RPP29 to RPP21 involves binding-coupled folding and stabilization of interfacial structures in RPP29.

When bound to its partner, RPP21 adopts the same overall L-shaped structure observed in the free state: a long arm containing the two N-terminal α-helices, a short-arm made up of the C-terminal β-sheet comprising the zinc ribbon, and a central linker connecting the two domains. However, in the complex, helix α1 of RPP21 extends through residues 9-17, indicating that binding is associated with induced fit in RPP21 as well. The RPP29 binding interface of RPP21 is clearly mapped to one face of the helix bundle by the cluster of residues identified with the largest CSPs. Backbone atoms of free and RPP29-bound RPP21 superpose with an overall RMSD of 2.5 Å. However,
when superposing the N-terminal helices and C-terminal sheet individually, the agreement is much better, with RMSD of 1.2 and 1.1 Å, respectively, reflecting the relatively poorly defined central linker in free RPP21 and corresponding uncertainty in interdomain orientation.[51] The N-terminal helices of *Pfu* RPP21 that form the binding interface adopt the same overall structure as observed in crystals of the *Pho* RPP21-RPP29 complex, with a backbone RMSD of 0.7 Å.

**Interface between RPP29 and RPP21**

The structure of the *Pfu* RPP21-RPP29 intermolecular interface identified from CSPs (Fig. 2.6) was defined by 284 unique intermolecular NOEs, many of which arise between methyl-bearing and aromatic residues (Appendix A). These NOEs highlight an extensive interface that buries approximately 2400 Å² of surface on both proteins (RPP29, 1100 Å²; RPP21, 1300 Å²) and is composed of hydrophobic, polar, and ionic interactions between residues provided in three separate structural elements: (1) the N-terminal region of RPP29 extends in an antiparallel fashion along RPP21 helix α₁; (2) RPP29 β₂ interacts with both helices of RPP21 in the center of the interface; and (3) the C-terminal helix of RPP29 stabilizes the end of RPP21 helix α₂ (Fig. 2.10).

In addition to hydrophobic interactions, conserved polar contacts likely play an important role in stabilizing the complex. In an unusual feature, I7₁^{RPP29} in strand β₂ adopts backbone φ / ψ torsion angles (140 ± 3 and -53 ± 12, respectively) that allow side chains of both I7₁^{RPP29} and D7₂^{RPP29} to participate in the RPP21 interface, through hydrophobic interactions and an intermolecular salt bridge, respectively. Moreover, this unusual backbone configuration orients the backbone carbonyl of I7₁^{RPP29} into the interface, where it is available for a stabilizing interaction with the side-chain hydroxyl of
Although intermolecular NOEs were observed between I71\textsubscript{RPP29} and Y39\textsubscript{RPP21}, these NOEs were not sufficient to constrain the side chain of Y39\textsubscript{RPP21} such that I71\textsubscript{RPP29} O and Y39\textsubscript{RPP21} O\textsuperscript{η} are in hydrogen-bonding distance in each member of the ensemble (3.3 ± 0.4 Å). This interaction was also observed in the crystal structure of the Pho RPP21-RPP29 complex,[97] and mutation of Y39\textsubscript{RPP21} to alanine was shown to be strongly deleterious to activity in an \textit{in vitro} reconstitution assay. Given the sequence conservation in this region of both protein, this is likely to be a conserved interaction across archaea and eukarya. Adjacent to this interaction, prominent intermolecular salt bridges appear to be formed between E47\textsubscript{RPP29} and R17\textsubscript{RPP21} and between D72\textsubscript{RPP29} and R38\textsubscript{RPP21} (Fig. 2.11). These charge pairs are highly conserved in the thermophilic and hyperthermophilic RPPs in archaea (Fig. 2.11 and 2.12). Many of the residues in the binding surfaces of both proteins are highly conserved or invariant (Fig. 2.11 and 2.12), suggesting a conserved binding interface between RPP29 and RPP21 in archaeal and eukaryotic RNase P.

\textbf{RNA binding surface}

RNA binding studies of \textit{Mth} and \textit{Saccharomyces cerevisiae} RPPs using the yeast three-hybrid assay showed that RPP29 binds the RPR in \textit{Mth} and yeast RNase P, and both RPP21 and RPP29 interact directly with H1 RNA (RPR) in human RNase P.[40; 41] Given that RPP29 and RPP21 form a tight and intimate complex, it is reasonable to imagine that this complex forms prior to assembly with the RPR and, thus, may constitute an RNA binding "unit" with one or more contiguous RNA binding surfaces. The electrostatic potential map of the \textit{Pfu} RPP29-RPP21 complex identifies two electropositive surface patches on different faces of the complex (Fig. 2.13). The larger
of these two surfaces (site 1) indeed spans both proteins, including the RPP21 central linker and helix $\alpha_5$ at the N-terminus, as well as $\beta_6$ and the C-terminal $\alpha$-helix of RPP29. Twenty highly conserved Arg and Lys residues are located on this surface patch, among them, R116, R120, and K123 in the C-terminal helix of RPP29 and K51, K53, K59, R60, R61, and K64 in the central linker of RPP21. The smaller site 2 is localized to the RPP21 $\beta$-sheet, with R77, R79, R81, K83, R84, K91, R100, and two His (H87 and H97), all extending on one face of the $\beta$-sheet. Of these, R100 is invariant and R77, R79, R81, and K91 are highly conserved in archaea and eukarya (Fig. 2.13). This analysis draws attention to two distinct faces of the binary complex as potential RNA binding interfaces for the RPR catalyst and pre-tRNA substrate.

**Footprinting**

After the potential RNA binding interface was proposed on the RPP21-RPP29 complex based on the electropositive patch, the information of the RPPs binding surface on RPR is crucial in order to gain the knowledge of the spatial organization of all the components in RNase P, which will lead to better understanding of the RNase P mechanism. The Gopalan Lab previously employed footprinting assays and demonstrated that the POP5-RPP30 binary complex, and not RPP21-RPP29, decreased the susceptibility to RNase T1 cleavage of a Pfu RPR deletion derivative containing only the C-domain.[58] As documented previously, technical problems complicated footprinting experiments with the full-length Pfu RPR. Therefore, Mja RNase P was explored as an alternative to it Pfu relative. Compared to Pfu RPR, Mja RPR is shorter, featuring a missing P8 and smaller P12 in the S-domain, and smaller P15/P16 in the C-domain. However, the information obtained from the footprinting experiments conducted
on *Mja* RPR and its cognate RPPs can be reasonably extrapolated into other archaeal RNase P due to the resemblance of the predicted secondary structures. All the enzymatic footprinting experiments presented here were performed by Dr. Dileep Pulukkunat (Gopalan Laboratory). Full-length *Mja* RPR was digested with either RNase T1 (which cleaves 3' to unpaired guanines) or RNase V1 (which cleaves base-paired nucleotides in RNA). In the presence of the *Mja* RPP21-RPP29 complex, paired regions P9, P10/11, and P12 in the S-domain are protected from V1 cleavage, while no protection by RPP21-RPP29 complex was observed in the C-domain (Fig. 2.14). This is consistent with the finding that addition of RPP21-RPP29 has effects on the full-length *Pfu* and *Mja* RPRs in terms of the catalytic rate and the NH\(^{4+}/Mg^{2+}\) requirement, but not the C-domain derivatives of archaeal RPRs, unlike POP5-RPP30. In contrast, *Mja* POP5-RPP30 protects from RNase T1 cleavage the RPR's C-domain, especially at or surrounding the universally conserved nucleotides (Fig. 2.14). The presence of all four RPPs does not show additional protection compared to the aggregate of the footprinting of each binary pair. Collectively, these data indicate that RPP21-RPP29 and POP5-RPP30 exclusively interact with the S and C domains, respectively, regardless of whether the other pair is present. Together with the enzymatic footprinting experiments performed in *Pfu* (Type A) and *Mja* (Type M) RNase P, we can reasonably expect similar RPR-RPPs interactions in archaeal and eukaryotic RNase P holoenzyme.

**CONCLUSION**

The NMR data presented here indicate that formation of the *Pfu* RPP29-RPP21 complex involves binding-coupled folding of structural elements in both proteins. The RPP29-RPP21 complex is defined by a combination of conserved interfacial hydrophobic and polar residues, including apparent salt bridges. Outside of the
structured cores, residues in the extreme N- and C-termini of both proteins indicate that these remain unstructured in the binary complex. Of particular interest are the long N-terminus of RPP29 (17 residues) and the C-terminus of RPP21 (13 residues), both of which possess an abundance of conserved and basic residues (Lys, Arg) that suggest a role in RNA binding. The surface of the protein-protein complex features two well-defined regions of positive electrostatic potential, highlighting possible RNA binding regions.

The structure of the *Pfu* RPP21-RPP29 complex, together with enzymatic footprinting data, permits informed speculation about the nature of the protein-RNA complex. Enzymatic footprinting of *Mja* RPP21-RPP29 on its cognate RPR indicates that this protein complex interacts only with the S domain of the RPR, consistent with earlier studies where it was shown to not change the rate of the phosphodiester bond-breaking step (Fig. 2.14).[44] Because of the lack of a 3D model of any archaeal RPR to date, the crystal structure of the bacterial RPRs serve as the best frame of reference. The crystal structure of the *T. thermophilus* (Type A bacterium) RPR S domain highlights intramolecular RNA-RNA interactions (especially between stems P13 and P12 and between P14 and P8) that stabilize the tertiary structure necessary for RPR-alone catalysis.[28] However, when compared to the bacterial type A RPR, P13 and P14 are absent from the S domain of *Pfu* and *Mja* RPRs, which instead exhibit an extended P12. It is tempting to propose that in archaeal RNase P, the 3D fold of the RPR is maintained and that the stabilizing RNA-RNA interactions in the bacterial enzyme are replaced by RNA-protein interactions in archaeal RNase P.[45] Thus, given the location of RNA binding by the RPP21-RPP29 complex, the function of this complex might be to compensate for the absence of the P13/P14 structural elements by mediating tertiary contacts between different parts of the RPR.
The presence of a secondary potential RNA-binding surface localized to the RPP21 \( \beta \)-sheet suggests a function unrelated to stabilizing RPR tertiary structure. Since human RPP21 has been reported to bind pre-tRNA,[98] it is conceivable that archaeal RPP21 might also be involved in substrate recognition. The S domain in bacterial RPR has been shown to recognize the T stem-loop (TSL) region in the ptRNA. The TSL-S domain interaction triggers a conformational change that aids catalysis by positioning the chemical groups and catalytically important Mg\( ^{2+} \) near the cleavage site in the C domain.[28; 99; 100] Thus, binding of RPP21-RPP29 to the archaeal RPR’s S domain might not only promote intra- and / or interdomain RPR cooperation but also directly (or indirectly via the RPR) mediate recognition of the TSL in the ptRNA. Further experimentation is required to evaluate if, indeed, the two potential RNA-binding sites identified in the electrostatic potential map of the RPP21-RPP29 complex play distinctive or overlapping roles in RPR and ptRNA binding.

In summary, we have determined the solution structure of the \( Pfu \) RPP29-RPP21 complex and found that poorly structured elements in each protein fold to form the intermolecular interface in the complex. Analysis of the electrostatic potential surface of this complex revealed two potential RNA-binding surfaces, the larger of which is composed of surface elements from both proteins. Finally, we have localized the RPP29-RPP21 complex to the S domain of archaeal RPR by enzymatic footprinting. These findings provide valuable new insights into mechanisms of RNP assembly and serve as important step towards a 3D model of this ancient RNP enzyme.
Figure 2.1. Interactions in archaeal RNase P with ptRNA. Yeast two- and three-hybrid assays have shown that extensive network exists in the archaeal RNase P, among which strong interactions are observed between RPP21 and RPP29, and between RPP30 and POP5. In addition, RPP21 and RPP29 are also involved in ptRNA substrate binding.
Figure 2.2. Structures of single protein subunits in archaeal RNase P. The secondary structure of *Pfu* RPR is from RNase P database.[26] The structures shown here are determined from *Pyrococcus horikoshii* RPP29 (Ph1771p, 1V76), RPP30 (Ph1877p, 1V77) and RPP38 (2CZW), *Pyrococcus furiosus* RPP21 (2K3R) and POP5 (2AV5). Secondary structures are indicated in colors (red: α-helix; blue: β-strand). The size of each component is not drawn in scale. The illustration does not reflect the spatial relationship between the RPPs and RPR.
Figure 2.3. Crystal structure of Pho RPP30-POP5. RPP30 and POP5 form a heterotetramer, composed of two copies of each proteins (PDB ID: 2CZV). On each side of the line is the RPP30-POP5 heterodimer.
Figure 2.4. Temperature effect on the spectra of Pfu RPP29. At 25°C, the spectrum (left) exhibits only a handful peaks, most likely due to the non-specific aggregation. Dramatic improvement is observed in the spectrum collected at 55°C (right), indicative of a well-folded protein.
Figure 2.5. Backbone assignments of free and bound *Pfu* RPP29 and RPP21. The residue numbers are shown in red.
Figure 2.6. Binding-coupled folding in *Pfu* RPP21 and RPP29 as detected by NMR. Overlay of 15N HSQC spectra of each 15N-labeled protein in the absence (black) and presence (red) of its partner illustrates the site-specific CSPs. Weighted average CSP values (c and d) are mapped onto cartoon diagrams of the proteins (e, RPP21 and f, RPP29) using a linear color ramp from gray (no change) to red (maximal CSP). Blue indicates residues whose signals were only observed in the spectra of the complex. Black, undetermined CSP.
Figure 2.7. Representative strips from $^{13}$C-filtered/edited NOE spectra. The two experiments were recorded on (a) [U-$^{13}$C, $^{15}$N]RPP21 bound to unlabeled Pfu RPP29 and (b) [U-$^{13}$C, $^{15}$N]RPP29 with unlabeled RPP21. The black (positive) cross peaks arise from intermolecular NOEs, while the red (negative) peaks result from incomplete suppression of self- and intramolecular NOEs. The $^1$H and $^{13}$C shifts of the labeled partner are indicated above and below each strip, while the y-axis corresponds to the shift of $^1$Hs from the unlabeled partner to which the labeled proton NOEs.
Figure 2.8. Solution structures of the Pfu RPP29-RPP21 complex. Ensemble of 10 lowest-energy structures superimposed on the backbone heavy atoms of residues 18-122 of RPP29 and residues 10-54, 57-81, and 88-104 of RPP21. Pfu RPP29 and RPP21 are labeled in red and blue, respectively.
Figure 2.9. Heteronuclear \( {}^1\text{H}-{ }^{15}\text{N} \) NOE data of \( Pfu \) RPP29 in complex with RPP21. Even in the presence of \( Pfu \) RPP21, the N-terminus (residues 1-16, highlighted in cyan) remains flexible, indicating this segment is not involved in binding to RPP21.
Figure 2.10. Interface of the Pfu RPP29-RPP21 complex. Electropositivity and electronegativity on the surface are shown in red and blue, respectively. Electrostatic potential maps on the surface of RPP29 (a) and RPP21 (b) illustrate that the interface is dominated by hydrophobic interactions surrounding prominent charge-charge interactions; the binding partner is shown as a ribbon (RPP21 in cyan and RPP29 in green). (c-e) Close-up showing three elements that stabilize the protein-protein complex. The lowest-energy structure is chosen as the representative.
Figure 2.11. Sequence alignment of select RPP21 homologs from Archaea and Eukarya. The alignment was generated with CLUSTALW, and illustrated using ESPRIPT2.2, in which red letters indicate a global similarity score of 0.7, and red boxed letters indicate invariant residues. Secondary structural elements represented in cartoon are observed in the NMR ensemble of Pfu RPP21 in complex with Pfu RPP29. Aligned sequences are from Pyrococcus furiosus (NCBI entry NP_579342), Pyrococcus horikoshii (NP_143456), Pyrococcus abyssi (NP_126253), Methanobacterium thermoautotrophicum (NP_276730), Methanosarcina barkeri (NCBI_entry YP_304815), Halobacterium sp. (NP_279631), Thermoplasma acidophilum (NP_393654), Methanococcus jannaschii (NP_247957), Methanococcus marapaludis (NP_079115), Methanococcus vannielii (YP_001322736), Archaeoglobus fulgidus (NP_068950), Saccharomyces cerevisiae (NP_012280), Schizosaccharomyces pombe (NP_596472) and Homo sapiens (NP_079115).
Figure 2.12. Sequence alignment of select RPP29 homologues from Archaea and Eukarya. The alignment was generated with CLUSTALW, and illustrated using ESPRIPT2.2, in which red letters indicate a global similarity score of 0.7, and red boxed letters indicate invariant residues. Secondary structural elements represented in cartoon are observed in the NMR ensemble of *Pfu* RPP29 in complex with *Pfu* RPP21. The N-terminal deletion derivatives subjected to studies described in Chapter 4 are illustrated on the sequence. Aligned sequences are from *Pyrococcus furiosus* (NCBI entry NP_579545), *Pyrococcus horikoshii* (NP_143607), *Pyrococcus abyssi* (NP_126024), *Methanobacterium thermoautotrophicum* (10QK_A), *Methanosarcina barkeri* (YP_303669), *Halobacterium sp.* (NP_280464), *Thermoplasma acidophilum* (NP_394719), *Methanococcus jannaschii* (NP_247455), *Methanococcus maripaludis* (YP_001549311), *Methanococcus vannielii* (YP_001323236), *Archaeoglobus fulgidus* (1TSS_A), *Saccharomyces cerevisiae* (NP_009816), *Schizosaccharomyces pombe* (NP_588479) and *Homo sapiens* (NP_006618).
Figure 2.13. Electrostatic potential map suggests two RNA binding regions. (a) Electrostatic potential map of the *Pfu* RPP29-RPP21 complex reveals two extended electropositive surface patches. The larger of these surfaces is composed of residues contributed by both proteins; the smaller is localized solely to the RPP21 zinc ribbon. (b) Highly conserved basic residues are shown on the ribbon diagrams of *Pfu* RPP29 (red) and *Pfu* RPP21 (green). The orientation is as in (a).
Fig. 2.14. Footprinting using RNase V1 and RNase T1 to identify RPP-binding sites in Mja RPR. Mja RPR labeled at the 5'-end (a) or 3'-end (b) was incubated either without (lanes 1, 3, 5, 7, and 9) or with (lanes 2, 4, 6, 8, and 10) RNase V1 (a) or RNase T1 (b). Mja RPR was present alone (lanes 1, 2, 7, and 8), with RPP21–RPP29 (lanes 3 and 4), with RPP30–POP5 (lanes 5 and 6), or with both binary complexes (lanes 9 and 10). Since reconstitution of the RPR with each binary RPP complex is performed in a buffer different from that used for reconstitution with both binary complexes together, two different control RNase T1/V1 digestions of the RPR are shown (lanes 1 and 2 for binary RPPs, and lanes 7 and 8 for both binary pairs). “Alk” and “T1” represent molecular size ladders generated by subjecting end-labeled, denatured Mja RPRs to alkaline hydrolysis and partial RNase T1 digestion, respectively. The RNase T1 cleavage sites were also mapped by using primer extension assays (data not shown).[44] (c) Summary of the RPP footprinting data depicted on a secondary-structure model of Mja RPR. Circled and boxed nucleotides indicate protection to RNase T1 and RNase V1, respectively; blue and red colors indicate regions of protection by RPP30–POP5 and RPP21–RPP29, respectively. The green arrow indicates an RPR position that showed increased susceptibility to RNase T1 in the presence of either RPP30–POP5 or all four RPPs. RNase V1 cleavages around nucleotides 130–150 suggest that the secondary structure as drawn may need to be revised. All the enzymatic footprinting experiments were performed by Dr. Dileep Pulukkunat (Gopalan lab).
Table 2.1. Structural statistics for the wild type *Pfu* RPP29-RPP21 complex.

<table>
<thead>
<tr>
<th></th>
<th>RPP29</th>
<th>RPP21</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NMR constraints</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOEs</td>
<td>2038</td>
<td>1407</td>
</tr>
<tr>
<td>Intraresidue ((i - j = 0))</td>
<td>833</td>
<td>606</td>
</tr>
<tr>
<td>Sequential ((i - j = 1))</td>
<td>475</td>
<td>364</td>
</tr>
<tr>
<td>Short range ((1 &lt; i - j &lt; 5))</td>
<td>204</td>
<td>203</td>
</tr>
<tr>
<td>Long range ((i - j &gt; 5))</td>
<td>526</td>
<td>234</td>
</tr>
<tr>
<td>Intermolecular (RPP29-RPP21)</td>
<td></td>
<td>472</td>
</tr>
<tr>
<td>Ambiguous</td>
<td>376</td>
<td>328</td>
</tr>
<tr>
<td>Hydrogen bonds(^a)</td>
<td>38</td>
<td>80</td>
</tr>
<tr>
<td>Dihedral angles</td>
<td>188</td>
<td>162</td>
</tr>
<tr>
<td><strong>Structure statistics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Violations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance violations &gt; 0.5 Å</td>
<td>1.50 ± 1.02</td>
<td></td>
</tr>
<tr>
<td>Dihedral angle violations &gt; 5°</td>
<td>1.63 ± 0.10</td>
<td></td>
</tr>
<tr>
<td>Deviation from idealized geometry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bonds (Å)</td>
<td>0.0046 ± 0.00008</td>
<td></td>
</tr>
<tr>
<td>Angles (°)</td>
<td>0.77 ± 0.02</td>
<td></td>
</tr>
<tr>
<td>Impropers (°)</td>
<td>0.56 ± 0.01</td>
<td></td>
</tr>
<tr>
<td>Ramachandran statistics (%)(^b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored</td>
<td>77.0</td>
<td></td>
</tr>
<tr>
<td>Additional allowed</td>
<td>21.1</td>
<td></td>
</tr>
<tr>
<td>Generously allowed</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>Disallowed</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Precision (RMSD from the mean structure)(^c)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Backbone atoms (Å)</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>All heavy atoms (Å)</td>
<td>0.87</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Hydrogen bonds were applied as upper-bound restraints between amide proton and oxygen atoms and between amide nitrogen and oxygen atoms.

\(^b\) Ramachandran analysis was performed using PROCHECK-NMR.[84]

\(^c\) Structure statistics were calculated using the 10 lowest-energy structures; RMSDs were calculated by superimposing residues 18-122 of RPP29 and residues 9-54, 57-81, and 86-104 of RPP21.
CHAPTER 3

STRUCTURE DETERMINATION OF THE PROTEIN COMPLEX OF

_Pfu_ VARIANT RPP29Δ17-RPP21V14 IN SOLUTION

INTRODUCTION

Before structure determination of the wild type _Pfu_ RPP21-RPP29 complex, we first solved the structure of the _Pfu_ RPP29Δ17-RPP21V14 complex by NMR spectroscopy. Initially, an RPP21 Ala14Val variant (denoted as RPP21V14) was used in our structural studies due to inadvertent selection of a clone with this spurious mutation. A few reasons prevented us from noticing the presence of this mutation. First, the 2D spectrum of free _Pfu_ RPP21 does not contain the signals from residues 1-17, even in the spectrum of the wild type protein. Second, the reconstituted RNase P from RPP21V14 showed a comparable activity as reported on the reconstituted RNase P from RPP21WT. Last but not least, the 2D spectrum of U-[¹⁵N]-RPP21V14 in the presence of unlabeled RPP29 indicates that RPP21V14 is capable of binding to RPP29, and forms a tight complex with RPP29.

Based on the RPP21 sequence alignment, Ala14 is highly conserved in the RPP21 family (Fig. 3.1b). It is located in the middle of the helix α₁ as observed in the crystal structure of _Pho_ RPP21.[52] This helical region (residues 1-16 in _Pfu_ and residues 4-21 in _Pho_) is unstructured in solution.[51] Ala is the amino acid that best promotes helical
Compared to Ala, Val is less favored in the middle of a helix ($\Delta\Delta G=0.61$ kcal/mol for Val under assumptions of $\Delta\Delta G=0$ kcal/mol for Ala and $\Delta\Delta G=1$ kcal/mol for Gly),[101] but is more favored in the $\beta$-sheet conformation.[102] Mutation of A14V may prevent the formation of a stable $\alpha$ helix of residues 1-16 in the free protein. The high conservation of A14 through RPP21 family also suggests an important role in 1) maintaining the integrity of the overall structure of RPP21, 2) binding to RPP29, and/or 3) participating in enzymatic function. However, ignorant of the RPP21 mutation, we made the complex from U-[13C,15N]-RPP29WT and unlabeled RPP21V14. Though we obtained high quality 2D HSQC spectra, almost no signal was detected in 3D spectra. The solution structures of the free proteins indicate that both RPP21 and RPP29 have flexible termini that are not observed in the NMR spectra.[48; 103] The presence of these flexible regions may result in the complex tumbling slower than what would be expected based on its size. This is associated with the faster $T_2$ relaxation, leading to quicker signal decay before detection after a long evolution time in 3D spectrum. Previously, the Pho RPP29 crystal structure was solved on an N-terminal deletion construct after failure of crystallization of the wild type protein.[48] The smaller construct renders a diffractable crystal, and is used to solve the structure at 2.0 Å. We pursued the same approach to circumvent the problems caused by the long termini, seeking to identify the minimal domain required for RPP21-RPP29 complex formation. Hypothetically, the smaller complex would rescue the signal loss in NMR spectra, but still accurately represent the wild type complex. Given that Pfu RPP29 has an extremely long N-terminus compared to the rest of the RPP29 family (Fig. 3.1a), we constructed several N-terminal deletion mutants of RPP29 for NMR structural studies, including $\Delta 5$, $\Delta 13$, $\Delta 17$, $\Delta 24$ and $\Delta 31$, where $\Delta X$ indicates the first $X$ residues are removed by
mutagenesis (Fig. 3.1a). After examining their binding ability to RPP21V14 and their effect on RNase P activity, RPP29Δ17 was chosen for use in the structure determination of the binary complex. The characterizations of RPP29WT, RPP29Δ17 and RPP29Δ24 were repeated using RPP21WT. Detailed characterization of all the mutants will be presented in Chapter 4.

NMR derived structure determination can be hindered by severe spectral overlap and broad linewidths resulting from the increased size of the molecules and the associated slower tumbling, respectively. Several techniques have been introduced to overcome these problems, such as deuterium labeling in order to minimize spin-spin relaxation[64; 65] and TROSY.[63] However, 13C-bound proton signals can not be detected in a fully-deuterated protein, limiting the information available for detailed structural analysis.

An alternative labeling strategy is to selectively protonate the specific positions, such as methyl groups of Ile, Leu and Val in a perdeuterated sample.[67; 104] This allows collection of enough distance restraints for structural analysis, while reducing spin-spin relaxation. ILV-protonated, U-[2H, 15N, 13C]-protein can be expressed by growing cells in 100% D2O minimal media with 15NH4Cl and 13C2H-glucose as the sole sources of nitrogen and carbon, followed by addition of U-[13C2H]-methyl-protonated Ile precursor (α-ketobutyrate) and Leu/Val precursor (α-ketoisovalerate) one hour prior to induction.[105] This labeling scheme has been successfully applied to an 82-kDa protein and has increased the upper bounds of practical NMR structure determination.[68]

We applied this method to simplify the NOESY spectra and analysis. The NOESY spectra of a 30-kDa protein-protein complex are expected to be crowded with thousands of NOE crosspeaks, which renders the unambiguous assignments of the NOEs nearly
impossible. As such, first an ILV methyl-protonated, U-[\(^2\)H, \(^{15}\)N, \(^{13}\)C]-RPP29Δ17 was prepared to form the complex with the unlabeled RPP21V14, thus "less-crowded" the NOESY spectra which only record the through space magnetization transfer to the amide and methyl protons on the ILV of RPP29Δ17. Furthermore, four 4D NOESY spectra were recorded on the complex such that the NOEs observed were only between the \(^{13}\)C-attached protons, or intramolecular NOEs of RPP29Δ17. They are HCCH-NOESY (NOEs between methyl protons in RPP29Δ17), HNNH-NOESY (NOEs between amide protons in RPP29Δ17), HN-N/CH NOESY (NOEs from amide protons or methyl protons to amide proton in RPP29Δ17), and HC-N/CH NOESY (NOEs from amide protons or methyl protons to methyl protons in RPP29Δ17).[106] These procedures greatly facilitated the structure determination of the RPP29Δ17-RPP21V14, especially in localization of the N-terminal region of RPP29Δ17.

Determination of the RPP29Δ17-RPP21V14 complex provided us with the first insight into this binary protein-protein complex, including detailed protein-protein binding interface and the putative RNA binding interfaces on the complex. When we were preparing the manuscript for the structure determination of the \(Pfu\) RPP29Δ17-RPP21V14 complex, the crystal structure of wild type \(Pho\) RPP21-RPP29 complex was published.[97] Despite the fact that the N-terminal helix of RPP21 (residues 9-17) is not observed in RPP21V14 in the complex, most of the features that characterize the RPP21-RPP29 binding interface and the electrostatic potential map of the entire complex are almost the same between \(Pho\) and \(Pfu\) RPP21-RPP29 complex. However, our NMR data indicate that a binding-coupled folding event occurs upon complex formation, which was not clear in the crystal structure. Moreover, the structure of RPP29Δ17-RPP21V14 provided a wealth of useful information for the subsequent
determination of the wild type RPP21-RPP29 complex.

MATERIALS AND METHODS

Protein preparation

_Pfu RPP29Δ17_

The gene encoding _Pfu RPP29Δ17_ was amplified by PCR using the full-length _Pfu_ RPP29/pET-33b plasmid[58] as the template and gene-specific DNA primers (forward: 5'-cc tcg acc ATG GGA TCG TAT CAA GAA ATT ATT GG-3'; reverse: 5'-cc taa ctc gag TCA TTT ACG CCA ACG C-3'. The lower-cased NTs not included in the _Pfu_ RPP29 gene and were included to enhance restriction enzyme activity; the underlined NTs are restriction enzyme sites of _Nco_ in the forward primer and _Xho_ in the reverse primer; the start and stop codons are indicated in bold) and then subcloned into the pET-33b vector between _Nco_ and _Xho_ sites. The sequence of _Pfu_ RPP29Δ17 was confirmed by sequencing. The U-[^15N]- and U-[^15N,^{13}C]-RPP29Δ17 protein was overexpressed and purified as described in Chapter 2.

The ILV-protonated, U-[^{2}H,^{15}N,^{13}C]-_Pfu RPP29Δ17_ (denoted as ILV-RPP29Δ17) was expressed as follows:[104] the _Pfu_ RPP29Δ17-pET33b plasmid was transformed into _E. coli_ BL21 (DE3) Rosetta cells and inoculated with 10 mL of LB. After ~4 hours, cells were harvested by centrifugation (6000xg, 10 min, room temperature) and resuspended in 50 mL of M9 minimal media in 99.8% D₂O (Sigma-Aldrich Inc.) with 1 g/L of ^15NH₄Cl and 2 g/L of ^13C,^{2}H-glucose as the sole nitrogen and carbon sources. Cells were harvested after overnight growth and resuspended in 1 L of the same media. Once the OD₆₀₀ reached 0.6, 0.1 g of 2-keto-3-(methyl-d₃)-butyric acid-1,2,3,4-^{13}C₄, 3-d₁, sodium salt (Isotec, Inc.) and 0.05 g of 2-ketobutyric acid-^{13}C₄,3,3-d₂ sodium salt (Isotec, Inc.)
were added as the precursors of Leu/Val and Ile, respectively. An hour later, 1 mM IPTG was added and cells were induced for 12 hours and harvested by centrifugation (6000xg, 10 min, room temperature). The ILV-RPP29Δ17 was purified as described for the wild type protein in Chapter 2.

*Pfu RPP21V14*

The overexpression and purification of *Pfu* RPP21V14 followed the same procedures described for the wild type protein in Chapter 2.

**NMR spectroscopy**

The *Pfu* RPP29Δ17-RPP21V14 complex was formed by combining one U-[¹⁵N]- or U-[¹⁵N,¹³C]-protein (~1 mM) with an unlabeled protein partner at ~1:1 molar ratio, with a slight excess of the unlabeled protein to ensure full saturation of the labeled protein.[51] Based on the observation that at room temperature both proteins aggregate, leading to very few crosspeaks and uneven peak intensities in the HSQC spectra, all data were recorded at 55°C on the same NMR spectrometers described in Chapter 2. NMR spectra were processed and analyzed with NMRPipe[91], NMRView[92] and CARA.[93]

Backbone assignments of free and RPP21V14-bound RPP29Δ17 were made by using standard triple resonance NMR spectra (HNCO, HNCA,CB, CBCA(CO)NH) recorded on 600 MHz.[72] Side chain assignments of RPP29Δ17 in complex with RPP21V14 were obtained from the following 3D spectra on 600 MHz: HBHA(CABCACO)NH, HCCH-correlated spectroscopy (COSY) (τₘ=12 ms) and HCCH-total correlated spectroscopy (TOCSY) (τₘ=12 ms).[72] The ¹H and ¹³C resonance assignments of methyl groups in ILV of bound RPP29Δ17 were further confirmed from
3D H(C)(CO)NH-TOCSY and C(CO)NH-TOCSY ($\tau_m$=12 ms, 600 MHz) spectra on ILV-RPP29Δ17 in the complex with unlabeled RPP21 (modified from standard Bruker pulse sequences by Dr. Chunhua Yuan at OSU CCIC). A $^{13}$C-edited aromatic NOESY spectrum ($\tau_m$=100 ms, 800 MHz) (courtesy of Dr. Justin Wu) was recorded for side-chain assignments of aromatic residues. Previously, Dr. Carlos D. Amero (Foster laboratory) has finished the backbone assignments of RPP21V14 in the absence and presence of RPP29WT using the standard 3D spectra (HNCO, HNCA, HNCACB, CBCA(CO)NH) recorded at 50°C.[51] These backbone assignments were easily transferred to the RPP29Δ17-bound RPP21V14 (at 55°C) with very small adjustment. Side chain $^1$H and $^{13}$C resonance assignments of RPP21V14 in the complex with RPP29Δ17 were accomplished using H(C)(CO)NH-TOCSY ($\tau_m$=12 ms) and C(CO)NH-TOCSY ($\tau_m$=12 ms) on 600 MHz.[72] Another $^{13}$C-edited aromatic NOESY spectrum ($\tau_m$=100 ms, 800 MHz) on the sample made from $^{15}$N,$^{13}$C-labeled RPP21V14 and unlabeled RPP29Δ17 served to complete the side chain assignments of aromatic residues in bound RPP21V14.

Interproton distance restraints were obtained from 3D $^{15}$N-separated NOESY-HSQC ($\tau_m$=100 ms), and $^{13}$C-separated NOESY-HSQC ($\tau_m$=100 ms) spectra recorded on 800 MHz on samples (one U-$[^{15}$N,$^{13}$C]-protein with unlabeled partner) dissolved in 10% and 99.8% D$_2$O, respectively. Intermolecular distance restraints were assigned initially from $^{13}$C-filtered/edited NOESY-HSQC spectra with $^1J_{CH}$ filter elements tuned to 120 and 160 Hz.[77; 78] ILV-RPP29Δ17 in complex with unlabeled RPP21V14 was used to record 4D $^{15}$N/$^{13}$C time-shared NOESY ($\tau_m$=100 ms, HC-N/CH NOESY),[106] and 3D $^{15}$N-edited NCH NOESY ($\tau_m$=100 ms) spectra,[107] the latter of which records the chemical shifts of amide nitrogen, methyl protons, and methyl carbons, while the NOEs transfer is from the methyl protons to the amide protons. Pulse sequences were modified/made by Dr.
Chunhua Yuan at OSU CCIC.

Chemical shift perturbations were calculated in the same manner as described in Chapter 2.

**Structure determination**

Distance restraints, backbone torsion angle restraints and hydrogen bonds were determined in the same fashion as described in Chapter 2.

Structure calculations were performed using simulated annealing protocols within the Xplor-NIH software suite.[83] Thirteen intramolecular NOEs of RPP29Δ17 were initially identified from $^{13}$C-separated NOESY-HSQC, $^{15}$N,$^{13}$C time-shared NOESY and $^{15}$N-edited NCH-NOESY spectra, which helped elucidate the interaction between the N-terminal region and the Sm-like core of RPP29Δ17. We used SANE to automatically assign the majority of the intramolecular NOEs from the well-folded core SANE[76] using chemical shifts and distance criteria based on the homology model generated by SWISS-MODEL[108; 109] using the crystal structure of Pho RPP29,[48] containing residues 36-127 only. Together with backbone torsion angle restraints and hydrogen bonds, an initial set of structures of RPP29Δ17 was generated and used as template for SANE, yielding additional NOEs introduced into the iterative structure calculations and resulting in more converged structure of RPP29Δ17. A similar strategy was applied for Pfu RPP21V14 using the solution structure of free Pfu RPP21 as the template.[51] Then, using the NOE and dihedral angle restraints from the refined structure ensembles of the individual proteins, and the 104 intermolecular NOEs identified from $^{13}$C-filtered/edited NOESY-HSQC spectra, an initial ensemble of the RPP29Δ17-RPP21V14 complex was generated from an extended template. An additional 137 intermolecular
NOEs were assigned by SANE using the initial docked ensemble of the complex as the template and introduced into the final refinement. A set of 10 lowest-energy structures was selected for analysis and evaluation with Xplor-NIH and PROCHECK-NMR (Table 3.2).[84]

**Isothermal titration calorimetry (ITC)**

The NMR, ITC and *in vitro* reconstitution experiments were initially performed unintentionally on *Pfu* RPP21V14. After realization of the A14V mutation of RPP21, the experiments were repeated using RPP21WT, RPP29WT and RPP29Δ17. Because the data provide a similar conclusion, we report here only the screening results using RPP21WT and RPP29WT and its N-terminal deletion mutants.

The proteins were purified as described in Chapter 2, and ≤ 2 mL of each protein was dialyzed against 0.7 L of ITC buffer (20 mM cacodylate, pH 6.7, 10 mM KCl, 0.3 mM ZnCl₂, 0.02% NaN₃) at least twice to ensure minimal buffer mismatch between the samples. Before each ITC run, the protein solutions were thoroughly vacuum degassed (5 min, experiment temperature). All ITC data were collected on a MicroCal VP-ITC at 55°C, with the cell containing RPP21 (~20 μM) and the syringe containing ~200 μM of RPP29 or RPP29Δ17. Each experiment consisted of an initial 3 μL injection followed by 5 μL injections, with a spacing of 400 s. The heat of dilution was calculated from the average of the last 10 injections. The exothermic heat pulse following each injection was integrated, adjusted for the heat of dilution, and fit to a one-site binding model in a nonlinear least squares manner.

\[
q_i = ν \Delta H[M] \left( \frac{K_A[L]_i}{1 + K_A[L]_i} - \frac{K_A[L]_{i-1}}{1 + K_A[L]_{i-1}} \right)
\]
where $q_i$ is the measured heat released or absorbed at the $i$th injection, $v$ is the known volume of the reaction, $[M]$ and $[L]$ are the concentration of the macromolecule and ligand at the $i$th injection, given by the injection volume and concentrations. The fitting procedure allowed determination of the stoichiometry ($n$), binding constant ($K_A$), enthalpy ($\Delta H$), and entropy ($\Delta S$) of the binding reaction.

**Reconstitution assays**

I-Ming Cho (a graduate student in Dr. Gopalan's lab) performed the kinetic studies on *Pfu* RNase P reconstituted from either RPP29WT or RPP29Δ17, to examine the effect of RPP29Δ17 on the RNase P activity in comparison of the RPP29WT. *In vitro* reconstitution assays were carried out under optimal conditions:[58] 50 mM Tris, pH 8.0, 120 mM MgCl$_2$, 100 mM NH$_4$OAc for partially (RPR + RPP21 + RPP29) reconstituted *Pfu* RNase P, and 50 mM Tris, pH 8.0, 30 mM MgCl$_2$, 800 mM NH$_4$OAc for fully (RPR + RPP30 + POP5 + RPP21 + RPP29) reconstituted *Pfu* RNase P. The assembly requires 10 min incubation at 37°C followed by another 10 min at 55°C. In the ptRNA processing reaction, either partially (0.5 μM of RPR and 1.25 μM of RPP21 and RPP29 derivatives) or fully (0.01 μM of RPR and 0.1 μM of each of 4 RPPs) reconstituted RNase P was mixed with internally radio-labeled *E. coli* ptRNA$^{Tyr}$ at final concentration ranging from 30 to 2000 nM. The initial velocities of substrate cleavage were measured based on the intensity of the product bands, and were subjected to Michaelis-Menten analyses using nonlinear least-squares fitting.

**Accession number**
NMR resonance assignments of *Pfu* RPP29Δ17-RPP21V14 have been deposited in the Biological Magnetic Resonance Data Bank (BMRB) with accession number 15935.

**RESULTS AND DISCUSSION**

**RPP29Δ17 is functionally comparable to the full-length RPP29**

The original objective of engineering *Pfu* RPP29Δ17 was to improve the quality of the 3D NMR spectra, which were almost entirely lost when using U-[15N,13C]-RPP29WT in complex with RPP21V14. Several RPP29 N-terminal deletion mutants were constructed as candidates for making the binary complex. Among them, RPP29Δ17 was the most promising. RPP29Δ17 adopts the same structured core as RPP29WT (residues 49-115), as indicated by the essentially identical 2D 15N-edited HSQC spectra of free RPP29Δ17 and RPP29WT, consistent with the previously reported solution structures of *Afu* [49] and *Mth* RPP29 (data shown in Chapter 4). [47] Upon binding to unlabeled RPP21, dramatic spectral changes were observed for RPP29Δ17, including shifted peaks and new signals, suggesting additional structural elements are stabilized through the interaction with RPP21 (Fig. 3.2). On the contrary, another RPP29 derivative, RPP29Δ24, which is devoid of the first 24 residues, loses almost all its affinity to bind RPP21. Under similar conditions, only a few peaks in the spectrum recorded in presence of RPP21 exhibit small CSPs, while the rest of the spectrum is largely unchanged. This indicates that RPP29Δ24 still possesses the same structured core in the free state, but loses its ability to bind RPP21 (data shown in Chapter 4). In addition, ITC experiments confirmed that RPP29Δ17 binds to RPP21 nearly as well as wild type protein (K_D of 0.352 μM for RPP29Δ17, and 0.250 μM for RPP29WT). The binding of RPP21 to RPPΔ24 is too weak to be detected by ITC under the same condition (Fig. 3.3
and Table 3.1). In addition to a similar affinity for RPP21, I-Ming Cho (Gopalan laboratory) verified that RPP29Δ17 is functionally comparable to the RPP29WT in \textit{in vitro} reconstitution assays. Michaelis-Menten analyses revealed that the $k_{\text{cat}}$ and $K_m$ values of the holoenzymes reconstituted with either RPP29WT or RPP29Δ17 are indistinguishable (Table 3.2). These observations indicate that RPP29Δ17 retains the structural features important for specific recognition of RPP21 and is thus a valid target for structure determination.

**NMR spectroscopy and structure determination of RPP29Δ17 in complex with RPP21V14**

Like the 	extit{Pfu} full-length protein, 2D HSQC spectra recorded on RPP29Δ17 in the absence and presence of RPP21V14 exhibit large chemical shift perturbations (CSP) in addition to the appearance of 32 new peaks in the bound spectrum (Fig. 3.4). The additional signals indicate that binding-induced folding of extra structural elements occurs in RPP29Δ17. The new peaks were assigned to the residues from the N- (residues 21-49) and C-termini (residues 116-122). Though the backbone atoms of residues 3-16 in RPP29WT are also assigned in the complex, $^1\text{H}-^1\text{H}$-heteronuclear NOE data indicate that this region is flexible. Correspondingly, very limited number of $^1\text{H}-^1\text{H}$ NOEs was observed from this region in the NOESY spectra, preventing us from determining its structure. We conclude that the first 16 residues in RPP29 are not involved in RPP21 binding and remain rather unstructured even in the presence of RPP21. Only residues 17-48 are stabilized in RPP29 by interaction with RPP21, which is predicted to form two short $\alpha$-helices (residues 26-32 and 40-48) by Chemical Shift Index (CSI) based on the chemical shifts (Fig. 3.5a).
With the application of Structure-assisted NOE assignments (SANE) to a homology model based on the crystal structure of Pho RPP29[48], the Sm-like core (residues 36-115) of RPP29Δ17 was readily determined and converged well. However, we were not able to use SANE to assign any NOEs from that region since it was not part of the homology model, and consequently the region was not constrained (residues 17-35). As indicated by CSPs, the binding interface of RPP29Δ17 is mainly composed of its two termini (Fig. 3.6). Therefore, defining the structure of the RPP29Δ17 N-terminus is of utmost importance. The 4D NOESY spectra collected on the ILV-RPP29Δ17 in complex with RPP21V14 contains many fewer peaks, which greatly simplified the data analysis (Fig. 3.7). Two crucial intramolecular NOEs to constrain this region were identified from 15N/13C, time-shared NOESY and 15N-edited NCH NOESY spectra recorded on RPP21V14-bound ILV-RPP29Δ17, which are from I24Hδ1 to G82Hδ and V86Hδ, respectively (Fig. 3.7a, strip 1). More NOEs were then assigned in the adjacent region, revealing spatial relationships between the N-terminal region and the hydrophobic core. These unique intramolecular NOEs, together with the short-range NOEs, hydrogen bonds and backbone dihedral angle restraints allowed convergence of the N-terminus of RPP29Δ17. The initial ensemble was then used as the template for SANE to assign more NOEs from the N-terminus, facilitating further refinement of the structure and orientation of the N-terminus, leading to the complete structure of RPP29Δ17 with good convergence. Binding to RPP21V14 triggers the formation of four short helices in RPP29Δ17, three in the N-terminus and one in the C-terminus (Fig. 3.6d). The helices are connected by extended loops, and wrap around the β-barrel core. The overall shape of the Sm-like core is unaltered compared to the free RPP29 (Fig. 3.8), except for a couple of residues with large CSPs, clustered on one face of the core (Fig. 3.6d). The N-
and C-terminal structural features are stabilized in the presence of RPP21V14, explaining the absence of their signals from the spectra of free RPP29.

**NMR spectroscopy and structure determination of RPP21V14 in the complex**

Amero *et al* have reported the solution structure of *Pfu* RPP21, which was also determined on the RPP21V14 mutant unintentionally[51]. The 2D HSQC spectra of free RPP21WT and RPP21V14 are almost identical (Fig. 3.9), indicating that free RPP21WT and RPP21V14 adopt the same structure. Thus, the A14V mutation evidently does not disrupt the solution structure of the free protein, and *Pfu* RPP21V14 accurately reflects the structure of the RPP21WT in the monomeric state.[51] The observation of RPP21’s unstructured N-terminal region (residue 4-17) in solution is in contrast to the *Pho* RPP21 crystal structure, where helix α₁ extends through residues 1-34[52]. However, no new amide signals appeared when U-[¹⁵N]-RPP21V14 is saturated with unlabeled RPP29Δ17 (or RPP29WT), unlike the nine new N-terminal resonances observed when RPP21WT binds to its partner (Fig. 3.5b and 3.9). Thus, the A14V mutation does impair the folding of the N-terminal helix in RPP21 upon RPP29 binding, though there is no effect on the RPP21V14 structure in the absence of its partner. Given the different conformational propensity between Ala and Val, this mutation might prevent the formation of the N-terminal α helix in RPP21V14 induced by interaction with RPP29. Moreover, the high conservation of A14 also suggests that it might be involved in RPP29 binding. Thus, the mutation might interfere with the arrangement of the binding interface between RPP21V14 and RPP29Δ17.
Using the NOE and backbone torsion angle restraints, the structure of RPP21V14 in complex with RPP29Δ17 was well determined for residues 17-104, the same region defined in free RPP21V14 (and in free RPP21WT based on nearly identical HSQC spectra)[51]. The interaction with RPP29Δ17 does not promote binding-coupled folding of residues 9-16 to stabilize into an α-helical conformation, as observed in RPP29-bound RPP21WT (Fig. 3.6). This confirms the importance of A14 of RPP21 in stabilizing the structure of RPP21 and interacting with RPP29. Otherwise, the overall structure of RPP21V14 in complex adopts the same L-shape structure, in good agreement with free RPP21V14. The superimposition of all the assigned residues (17-101 of RPP21V14) reports the RMSD of 2.12 Å. However, when aligning only the α-helices (residues 17-50) or the beta-sheet (residues 76-101), the RMSD are improved to 1.24 Å and 1.07 Å, indicating a poorly defined central linker region (Fig. 3.8).

Solution structure of the *Pfu* RPP29Δ17-RPP21V14 complex

217 intermolecular NOEs were used to refine the binding interface (Fig. 3.10), including 103 from $^{13}$C-filtered/edited NOESY and 114 from $^{13}$C-separated NOESY-HSQC automatically assigned with SANE. Since residues 9-16 of RPP21 remain unstructured in the complex and do not participate in RPP29V14 binding, the number of assigned intermolecular NOEs for RPP29Δ17-RPP21V14 (217) is understandably less than the one of RPP29-RPP21WT (472). The decreased number of intermolecular NOEs may also be attributed to the relatively loose packing between RPP29Δ17 and RPP21V14, especially around the area near V14. The convergence of the RPP29Δ17-RPP21V14 complex is also a little worse than that of the wild type complex, reported by the mean RMSD of 0.71 and 1.00 Å for backbone and heavy atoms, respectively (Fig.
Both RPP29WT and RPP29Δ17 are only well-defined of residues 18-123 in complex. The very limited number of NOEs observed from residues 3-16 did not allow us to restrain this segment into a stable conformation in RPP29WT. This suggests that the first 17 residues of RPP29 are not involved in RPP21 binding, and rationalizes that the RPP29Δ17-RPP21V14 can well represent the wild type complex.

Overall, the largest difference between RPP29Δ17-RPP21V14 and wild type RPP29-RPP21 lies in the unstructured N-terminal region (residues 9-16) in RPP21V14, which forms into an α helix in RPP21WT. Otherwise, the complexes share a very similar structure, with backbone RMSD of 1.94 Å, (RPP29 core residues 18-122 and RPP21 helix bundle residues 17-50) (Fig. 3.8). The composition of the binding interface in RPP29Δ17-RPP21V14 is in the same fashion as wild type complex, including 1) the N-terminal region of RPP29 interacts with RPP21V14 helix α1 in an antiparallel fashion, (though the latter is shorter than the RPP21WT helix α1); 2) the center of RPP21V14 helix bundle is stabilized by RPP29 β2; 3) RPP29 helix α4 at the C-terminus contacts the end of RPP21V14 helix α2. Unsurprisingly, the electrostatic potential maps of the binding interface on each protein, and of the surface on whole RPP29Δ17-RPP21V14 complex are essentially identical with those of RPP29-RPP21 wild type complex (Fig. 3.12). All residues contributing to form the protein-protein and potential protein-RNA interfaces, are located in the same positions and orientations observed in the wild type complex. The solution structures of RPP21-RPP29 complex (as well as the RPP21V14-RPP29Δ17 complex) are also in good agreement with the Pho RPP21-RPP29 crystal structure[97].
The structure of the wild type RPP29-RPP21 complex has illustrated that A14 has two significant structural roles, in accord with its high conservation in the RPP21 family (Fig. 3.13). One is to hold the two helices together through hydrophobic interactions between the A14 methyl group (helix $\alpha_1$) and the methyl groups of V46 and A50 (helix $\alpha_2$) when RPP29 is present. The other is to participate in RPP29 binding by hydrophobic interaction with L121$^{\text{RPP29}}$. Due to its preference for β-sheet conformation, Val may disrupt the middle of helix $\alpha_1$, and / or interfere with the methyl-methyl interaction with V46 and A50, which stabilizes the newly-formed part of helix $\alpha_1$. Moreover, the introduction of two methyl groups upon Ala to Val mutation in the tight, compact binding interface might introduce steric conflict, leading to at least a different local packing interface or possibly total unfolding of the region.

The loss of RPP29Δ24’s ability to bind RPP21 can also be explained now with the structure of the RPP29-RPP21 (or RPP29Δ17-RPP21V14) complex in hand. As mentioned above, the first two identified intramolecular NOEs to locate the N-terminal region of RPP29 are from the methyl group of I24 to amide protons of G82 and V86. Stabilized by the hydrophobic contacts of I24-V86 and I24-G82, the helix $\alpha_1$ is able to be appropriately oriented, allowing I23$^{\text{RPP29}}$ to interact with Y31$^{\text{RPP21}}$ and S32$^{\text{RPP21}}$ (Fig. 3.13). Removal of the first 24 residues deletes the entire helix $\alpha_1$ and abolishes the interaction between RPP29 helix $\alpha_1$ and RPP21 helix $\alpha_1$. Furthermore, the rest of the N-terminal region (helices $\alpha_2$ and $\alpha_3$) may no longer be anchored to the correct position/orientation, destroying other intermolecular interactions, including from I29$^{\text{RPP29}}$, A33$^{\text{RPP29}}$, and E47$^{\text{RPP29}}$. Sequence alignment in RPP29 family reveals that residues 1-34 are unique to Pyrococcus subgroup, not present in any other archaeal RPP29 (Fig. 3.1b), suggesting that this segment is dispensable. However, our RPP29 N-terminal deletion mutants clearly illustrate that Pfu RPP29's ability to bind RPP21 heavily relies on the residues 18-
24. In addition to the $\beta_2$ and the C-terminal $\alpha_4$, *Pyrococcus* RPP29 appears to have evolved to use the extra fragment at N-terminus to provide further stabilization in the binding interface; perhaps these additional interactions are stabilizing when living in harsh conditions near 100°C.

**CONCLUSION**

Despite our ignorance of the A14V mutation in RPP21 which complicated the NMR studies of the complex with full-length RPP29, we managed to circumvent the situation by successfully collecting high quality NMR spectra on the complex formed from RPP29 N-terminal deletion construct RPP29Δ17 and RPP21V14. 4D NMR spectra were recorded on the complex made from ILV-RPP29Δ17 and unlabeled RPP21 so as to simplify the spectra and the analysis (Fig. 3.7). Intermolecular NOEs from these experiments helped to define the structure of the N-terminal region (residues 17-35) of RPP29Δ17, which had not been previously observed. Compared to the solution structure of the wild type *Pfu* RPP21-RPP29 complex (Chapter 2), the structure of the RPP29Δ17-RPP21V14 complex is considered to be a very good representative of the wild type complex, which characterizes of the crucial structural elements for complex formation. This also supports the observations from *in vitro* reconstitution assays that RNase P formed from RPP21V14 and RPP21WT have comparable activities (Chapter 4). The solution structure of the RPP29Δ17-RPP21V14 complex is well-defined for the residues 18-123 of RPP29Δ17 and residues 17-54, 57-81 and 86-104 of RPP21V14. The only difference is the shorter helix $\alpha_1$ in RPP21V14, compared to the RPP21WT. Binding-coupled folding was observed in RPP29Δ17, but not in RPP21V14 presumably due to the disruption introduced by A14V mutation. Of the most importance, the structure of the RPP29Δ17-RPP21V14 complex reveals the details of binding interface between RPP29
and RPP21 for the first time, and provides invaluable information for the subsequent structure determination of wild type RPP29-RPP21 complex. In addition, it also provides structural insights into the biochemical assays performing using RPP29 N-terminal deletion mutants (see Chapter 4).
Figure 3.1. Sequence alignment of selected RPP29 and RPP21 homologs from Archaea and Eukarya. The alignment was generated with CLUSTALW, and illustrated using ESPRIPT2.2, in which red letters indicate a global similarity score of 0.7, and red boxed letters indicate invariant residues. Secondary structural elements represented in cartoon are observed in the crystal structures of Pho RPP29 (a) and RPP21 (b) in the complex. 

(a) Aligned sequences are from Pyrococcus furiosus (NCBI entry NP_579545), Pyrococcus horikoshii (NP_143607), Methanobacterium thermoautotrophicum (10QK_A), Methanococcus jannaschii (NP_247439), Archaeoglobus fulgidus (1TSF_A), Saccharomyces cerevisiae (NP_009816), Schizosaccharomyces pombe (NP_588479) and Homo sapiens (NP_006618). Black arrows show the N-terminal deletion mutants we constructed in Pfu RPP29. The light red arrow shows the Pho N-terminal deletion mutant Pho RPP29N31 used for crystallization, the electron density map of which starts at residue 36, indicated by the dark red arrow. The blue arrow illustrates the conserved core of Pfu RPP29, generated by limited trypsin proteolysis. (b) Aligned RPP21 sequences are from Pyrococcus furiosus (NCBI entry NP_579342), Pyrococcus horikoshii (NP_143456), Methanobacterium thermoautotrophicum (NP_276730), Methanococcus jannaschii (NP_247957), Archaeoglobus fulgidus (NP_068950), Saccharomyces cerevisiae (NP_012280), Schizosaccharomyces pombe (NP_596472) and Homo sapiens (NP_079115). Black arrow points the highly conserved A14 in the sequences.
Figure 3.2. Overlay of HSQC spectra of RPP29Δ17 and RPP21V14 in the absence and presence of its unlabeled partner. Large chemical shift perturbations were observed in both RPP29Δ17 (left) and RPP21V14 (right) spectra, but RPP21V14 bound spectrum does not contain new peaks, indicating no binding-coupled folding occurs in RPP21V14 upon interaction with RPP29Δ17.
Figure 3.3. Isothermal titration calorimetry of RPP21 binding to RPP29 derivatives. ITC experiments confirm that the binding affinities of RPP29WT and RPP29Δ17 to RPP21WT are comparable. Equilibrium dissociation constant, $K_d$, were 0.250 μM of RPP29WT, and 0.352 μM of RPP29Δ17 (Table 3.1). RPP29Δ24 shows no detectable binding under these conditions.
Figure 3.4. Backbone amide assignments of RPP29Δ17 and RPP21V14 in the complex. The NMR spectra were collected on the complex formed from Pfu RPP29Δ17 and RPP21V14 with one protein labeled and the other unlabeled. Labels are residue numbers. The unlabeled peaks are from side chains or unassigned.
Figure 3.5. Chemical Shift Index (CSI) of *Pfu* RPP29Δ17 and RPP21V14 in complex. The chemical shifts of the backbone atoms (H\(\text{N}\), H\(\alpha\), C\(\alpha\), C\(\beta\), C\(O\) and N) are compared to the chemical shifts of a random coil conformation. The \(\alpha\)-helical and \(\beta\)-strand propensities are shown in red and blue, respectively, while the random coil conformation is shown in gray. Once three consecutive residues share the same propensity, CSI predicts the secondary structure of the fragment (cartoon shown at the bottom of the sequences).
Figure 3.6. Chemical shift perturbations illustrate the binding interfaces on RPP21V14 and RPP29Δ17. Weighted average CSPs (a and b) are mapped onto cartoon diagrams of the proteins (c, RPP21V14 and d, RPP29Δ17) using a linear color ramp from gray (no CSP) to red (maximal CSP). Undetermined CSPs are black. Additional structural elements at both termini of RPP29Δ17 are formed upon binding, and highlighted in blue. No binding-induced folding was observed in RPP21V14.
Figure 3.7. Representative strips from 4D spectrum collected on the ILV-RPP29Δ17 in complex with the unlabeled RPP21V14. NOEs that can not be recorded in the 13C-edited NOESY spectrum (b) are observed in the 4D HN-N/CH NOESY spectrum (a) collected on the complex made from ILV-RPP29Δ17 and unlabeled RPP21V14. Only with help of the ILV-methyl labeled sample and 4D NMR spectrum, two crucial NOEs, V86H\(^N\) and G82H\(^N\) to I24H\(^\delta\) (strip 1 in a), were able to be identified and help to constrain the N-terminal region of RPP21Δ17 to its core.
Figure 3.8. Superimpositions of RPP29 and RPP21. Superimposition (a) of the helix bundle (residues 17-50) of free (gray) and bound (yellow) RPP21V14 indicates that the overall L-shape is unaltered upon binding. However, additional N-terminal α-helix is induced to form in the RPP21WT (cyan) by interacting with RPP29 (c), but not in the RPP21V14. Binding of RPP21 does not change the hydrophobic core (superimposed on residues 36-122) of RPP29 (b), but triggers the formation of the terminal structural elements.
Figure 3.9. Overlay of $^{15}$N HSQC spectra of RPP21V14 and RPP21WT in the free state. There is an identical number of peaks present in the RPP21WT spectrum (red) and in the RPP21V14 spectrum (black). Most of the peaks remain at the same position, except that only a handful peaks shift slightly. This suggests that RPP21WT and RPP21V14 share the almost same structure in the monomeric state, regardless of the A14V mutation.
Figure 3.10. Representative strips from $^{13}\text{C}$-filtered/edited NOE spectra collected on RPP29Δ17-RPP21V14 complex. The two experiments were recorded on (a) $[^{13}\text{C},^{15}\text{N}]RPP29\Delta17$ bound to unlabeled Pfu RPP21V14 and (b) $[^{13}\text{C},^{15}\text{N}]RPP21V14$ with unlabeled RPP29Δ17. The black (positive) cross peaks arise from intermolecular NOEs, while the red (negative) peaks result from incomplete suppression of self- and intramolecular NOEs. The $^1\text{H}$ and $^{13}\text{C}$ shifts of the labeled partner are indicated above and below each strip, while the $y$-axis corresponds to the shift of $^1\text{H}$s from the unlabeled partner to which the labeled proton NOEs.
Figure 3.11. Ensemble of the *Pfu* RPP29Δ17-RPP21V14 complex. Ten lowest-energy structures are superimposed on the backbone heavy atoms of residues 18-122 of RPP29Δ17 and residues 17-54, 57-81 and 86-101 of RPP21V14. *Pfu* RPP29Δ17 and RPP21V14 are labeled in pink and yellow, respectively. The secondary structures are highlighted.
Figure 3.12. Electrostatic potential map of the *Pfu* RPP29Δ17-RPP21V14 complex. The electropositivity and electronegativity are shown in blue and red, respectively. The protein-protein interface features central electrostatic contacts, surrounded by the hydrophobic interactions (a and b). Two putative RNA binding sites are in good agreement with wild type *Pfu* RPP29-RPP21 complex, with slightly different localization of the electropositive surface on site 1, which is most likely due to the different orientation of the basic side chains (c).
Figure 3.13. Structural details of A14\textsuperscript{RPP21} and I24\textsuperscript{RPP29}. The lowest-energy structure of the wild type complex (RPP21 in cyan and RPP29 in green) is used here. (a) Close-up showing that A14\textsuperscript{RPP21} is involved in intra- and inter-molecular interactions. (b) Intramolecular NOEs were observed between I24Hδ1 and HN's of G82 and V86. These interactions are responsible for locking down the N-terminal helix α1 to the Sm-like core in RPP29.

Table 3.1. Steady-state kinetics and thermodynamics of RPP29\textsubscript{WT} and RPP29\textsubscript{Δ17}.*

<table>
<thead>
<tr>
<th></th>
<th>RPP29\textsubscript{WT}</th>
<th>RPP29\textsubscript{Δ17}</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_\text{cat}$ (min\textsuperscript{-1})\textsuperscript{a}</td>
<td>9.4 ± 1.4</td>
<td>9.5 ± 1.7</td>
</tr>
<tr>
<td>$K_m$ (μM)\textsuperscript{a}</td>
<td>0.24 ± 0.06</td>
<td>0.27 ± 0.06</td>
</tr>
<tr>
<td>Stoichiometry (N)\textsuperscript{b}</td>
<td>1.001 ± 0.003</td>
<td>1.011 ± 0.003</td>
</tr>
<tr>
<td>$K_a$ (μM)\textsuperscript{b}</td>
<td>0.250 ± 0.016</td>
<td>0.352 ± 0.028</td>
</tr>
<tr>
<td>ΔG (kcal·mol\textsuperscript{-1})\textsuperscript{b}</td>
<td>-9.901 ± 0.646</td>
<td>-9.966 ± 0.795</td>
</tr>
<tr>
<td>ΔH (kcal·mol\textsuperscript{-1})\textsuperscript{b}</td>
<td>-29.67 ± 0.186</td>
<td>-24.66 ± 0.102</td>
</tr>
<tr>
<td>ΤΔS ((kcal·mol\textsuperscript{-1})\textsuperscript{b})</td>
<td>-19.77 ± 1.29</td>
<td>-14.69 ± 1.17</td>
</tr>
</tbody>
</table>

* The standard deviations are reported on the three experiments.

\textsuperscript{a} $k_\text{cat}$ and $K_m$ were measured by \textit{in vitro} reconstitution assays at the optimal condition for the fully-reconstituted RNase P (RPR+POP5+RPP30+RPP21+RPP29 derivatives).

\textsuperscript{b} Thermodynamic parameters were from ITC experiments performed at 55°C.
Table 3.2. Structural statistics for the *Pfu* RPP29Δ17-RPP21V14 complex.

<table>
<thead>
<tr>
<th>NMR constraints</th>
<th>RPP29Δ17</th>
<th>RPP21V14</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOEs</td>
<td>2139</td>
<td>1622</td>
</tr>
<tr>
<td>Intraresidue ((i - j = 0))</td>
<td>963</td>
<td>781</td>
</tr>
<tr>
<td>Sequential ((i - j = 1))</td>
<td>443</td>
<td>400</td>
</tr>
<tr>
<td>Short range ((1 &lt; i - j &lt; 5))</td>
<td>214</td>
<td>169</td>
</tr>
<tr>
<td>Long range ((i - j &gt; 5))</td>
<td>519</td>
<td>272</td>
</tr>
<tr>
<td>Intermolecular ((RPP29-RPP21))</td>
<td>217</td>
<td></td>
</tr>
<tr>
<td>Ambiguous</td>
<td>367</td>
<td>365</td>
</tr>
<tr>
<td>Hydrogen bonds(^a)</td>
<td>42</td>
<td>60</td>
</tr>
<tr>
<td>Dihedral angles</td>
<td>180</td>
<td>158</td>
</tr>
</tbody>
</table>

Structure statistics

<table>
<thead>
<tr>
<th>Violations</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance violations &gt; 0.5 Å</td>
<td>0.1 ± 0.3</td>
<td></td>
</tr>
<tr>
<td>Dihedral angle violations &gt; 5°</td>
<td>1.32 ± 0.05</td>
<td></td>
</tr>
</tbody>
</table>

Deviation from idealized geometry

| Bonds (Å)                        | 0.0032 ± 0.00005 |        |
| Angles (°)                       | 0.55 ± 0.01      |          |
| Impropers (°)                    | 0.39 ± 0.01      |          |

Ramachandran statistics (%)\(^b\)

| Favored                          | 80.9    |          |
| Additional allowed               | 17.9    |          |
| Generously allowed               | 0.6     |          |
| Disallowed                       | 0.6     |          |

Precision (RMSD from the mean structure)\(^c\)

| Backbone atoms (Å)               | 0.71    |          |
| All heavy atoms (Å)              | 1.00    |          |

\(^a\) Hydrogen bonds were applied as upper-bound restraints between amide proton and oxygen atoms and between amide nitrogen and oxygen atoms.

\(^b\) Ramachandran analysis was performed using PROCHECK-NMR.

\(^c\) Structure statistics were calculated using the 10 lowest-energy structures; RMSDs were calculated by superimposing residues 18-122 of RPP29Δ17 and residues 17-54, 57-81, and 86-101 of RPP21V14.
CHAPTER 4

CHARACTERIZATION OF *Pfu* RPP29 N-TERMINAL DELETION MUTANTS:

MINIMAL DOMAIN OF *Pfu* RPP29 FOR ASSEMBLY WITH *Pfu* RPP21

INTRODUCTION

Structure determination by NMR spectroscopy requires high quality spectra, characterized by 1) high S/N ratio, 2) wide dispersion of signals, and 3) narrow linewidths, all of which can be problematic for the application of NMR spectroscopy to large (> 25 kDa) macromolecular systems.[66; 110] First of all, large molecules have large surface area and tumble slowly in solution, resulting in faster relaxation of transverse magnetization (shorter T$_2$). In 3D and 4D spectra, a short T$_2$ becomes a limiting issue: rapidly decaying signals will not be detected after the multiple pulses and periods of evolution in these spectra. Moreover, large proteins intrinsically produce more signals in the spectra, leading to inevitable resonance overlap. Last but not least, shorter T$_2$ also leads to broad linewidth, which worsens the S/N ratio and contributes to spectral overlap. Isotopic labeling of certain types of amino acids, or of a fragment/domain has greatly helped to reduce the spectral overlap,[111] but this approach can be painstaking and is limited by the protein sequence. On the other hand, several breakthrough techniques have helped to overcome the fast relaxation related problems of large molecules. These include introduction of the TROSY (Transverse Relaxation Optimized...
Spectroscopy) pulse sequence, which retains only the slower relaxing component of a coupling multiplet,[63] uniform deuterium labeling (to reduce spin-spin interactions by diluting protons),[64; 65] and data collection at elevated temperature (to promote faster tumbling of molecules).[66] Initial 3D NMR spectra of the *Pfu* RPP29-RPP21V14 complex were of poor quality, even at 55°C. As a 30-kDa protein-protein complex, it approaches the limit of structure determination by NMR spectroscopy.[66] The A14V mutation in RPP21 worsens the situation by preventing proper coupled folding of RPP21 N-terminal region (residues 9-17) upon binding RPP29, and leading to an even longer unstructured terminus. Thus, although induced fit occurs in RPP29 upon binding, there are more parts that remain unstructured in the RPP29Δ17-RPP21V14 complex, than in the wild type complex. These extended flexible termini likely cause the complex to tumble slower, like a larger protein, and lead to a faster $T_2$ relaxation resulting in almost no signal in 3D NMR spectra. In addition, there is possibly meta-stable (transient) structure that leads to exchange broadening. Interestingly, high quality 3D NMR spectra could be recorded on wild type *Pfu* RPP29-RPP21 complex, the flexible termini of which are only slightly shorter (9 residues less). Evidently, the binding-coupled folding of N-terminus of RPP21 (residues 9-17) is very important for formation of a tight complex between RPP29 and RPP21.

At the time, we were not aware of the presence of the A14V mutation in RPP21 and did not have detailed knowledge of the structure of the RPP29-RPP21 complex. After TROSY-based 3D experiments could not recover adequate signals and before performing costly deuterium labeling, we explored use of truncated protein constructs which might be viable for complex formation while showing a reasonable S/N ratio in the
3D spectra. Of the two proteins, RPP29 offers more freedom for engineering new constructs because of its extremely long N-terminus (residues 1-35), unique to the Pyrococcus genus (Fig. 4.1). However, the structure of this segment was unclear because either the reported structures of the free protein were solved using the RPP29 variants that do not have this longer N-terminal extension (Mth[47] and Afu)[49] or the construct used for structure determination was engineered to remove the N-terminus to facilitate crystallization (Pho RPP29N31).[48] As such, it appears that this region is dispensable for the overall structural integrity of RPP29.

In order to identify the smallest construct of RPP29 that can be used to form the complex with RPP21 for structure determination, we engineered seven N-terminal deletion mutants of RPP29, and evaluated their binding to RPP21 by NMR spectroscopy and ITC, and their effect on RNase P activity by in vitro reconstitution assays. Our results suggest an important role for the N-terminus of Pfu RPP29 in RPP21 binding as well as RNase P activity. This work allowed us to choose Pfu RPP29Δ17 for assembly with RPP21 (RPP21V14), which later proved to be a good structural representative of the wild type complex.

We initially performed the NMR, ITC, and in vitro reconstitution experiments using Pfu RPP21V14 (by Hsin-Yue Tsai, the Gopalan lab), and later repeated the same experiments using Pfu RPP21WT to control for the potential variations introduced by A14V mutation of RPP21. Because results from Pfu RPP21WT and RPP21V14 arrive at the same conclusion, this chapter only reports the data using Pfu RPP21WT with RPP29 deletion mutants.

**MATERIALS AND METHODS**
Protein preparation

*Pfu* RPP29WT and N-terminal deletion mutants

Seven *Pfu* RPP29 mutants were constructed, with different length of deletions at the N-terminus, named RPP29Δ5, RPP29Δ13, RPP29Δ17, RPP29Δ24, RPP29Δ31, RPP29Δ36, and RPP29Δ42, based on the number of N-terminal residues removed. Each gene encoding the specific RPP29 N-terminal deletion mutant was amplified by PCR using the full-length *Pfu* RPP29/pET-33b plasmid[58] as the template and gene-specific DNA primers (Table 4.1) and then subjected to restriction enzyme digestion (*BamH*I and *Xho*I) to generate sticky ends. The pET-33b vector was also digested by the same set of restriction enzymes, then cleaned by the PCR Purification Kit (Qiagen). The digested PCR products and pET33b vector were mixed at molar ratio of 1:9, and ligated with T4 DNA Ligase (New England BioLabs). Then, the plasmid with each new construct was transformed into DH5α cells for plasmid production. The sequences of *Pfu* N-terminal deletion derivatives were confirmed by automated DNA sequencing. The unlabeled and U-[15N]-RPP29ΔX proteins were overexpressed and purified as described for the wild type protein in Chapter 2.

*Pfu* RPP21

The overexpression and purification of *Pfu* RPP21 followed the same procedures described in Chapter 2.

Limited proteolysis identified the folded core of RPP29
100 µg of purified *Pfu* RPP29WT was subjected to limited proteolysis by 1% trypsin (w/w) at room temperature for 1 hour. Then, the reaction was quenched by adding 5% PMSF (phenylmethanesulphonylfluoride) and stored on ice. The digested proteins were analyzed by electrospray mass spectroscopy (Q-TOF-II, Micromass).

**NMR spectroscopy**

For each U-[\textsuperscript{15}N]-*Pfu* RPP29ΔX mutant, a 2D HSQC spectrum was recorded on the free protein first, then another 2D spectrum was recorded on the complex made from the U-[\textsuperscript{15}N]-*Pfu* RPP29ΔX (~0.3 mM) and unlabeled *Pfu* RPP21, with a slight molar excess of unlabeled RPP21. We also performed the NMR assay on the complex with reverse labeling, that is with U-[\textsuperscript{15}N]-*Pfu* RPP21 and unlabeled *Pfu* RPP29ΔX. NMR spectra were processed and analyzed using NMRPipe[91] and NMRView.[92]

**Isothermal Titration Calorimetry (ITC)**

As documented in Chapter 3 Materials and Methods, all proteins were purified, and dialyzed into ITC buffer (20 mM Cacodylate, pH 6.7, 10 mM KCl, 0.3 mM ZnCl\textsubscript{2}, and 0.02% NaN\textsubscript{3}) twice. Typically, an ITC experiment is composed of a series of 5 µL injections of *Pfu* RPP29WT (or RPP29Δ17) at ~200 µM into *Pfu* RPP21WT (~20 µM) with an interval of 400 s between injections. Most experiments were performed in triplicate.

**Reconstitution assays**
The substrate cleavage kinetic studies were performed by graduate student I-Ming Cho in the Gopalan lab. RNase P was assembled either partially (RPR + RPP21 + RPP29) or fully (RPR + RPP30 + POP5 + RPP21 + RPP29), under their optimal conditions (See Chapter 3 for detailed buffer compositions and assembly steps). The final concentration of the substrate (E. coli ptRNA\textsubscript{Tyr}) was kept at 16 μM and 625 nM for partially (1 μM) and fully (10 nM) reconstituted RNase P. Thus, under conditions that the substrate is in great excess over enzyme, the initial velocities of substrate cleavage were extracted from the linear fit of product concentration vs. time, and then were converted to turnover number (number of substrate cleaved per enzyme per minute). The activities of RNase P reconstituted from RPP29 N-terminal mutants are reported as the “relative activity” towards the activity of wild type RNase P.

RESULTS AND DISCUSSION

Limited trypsin proteolysis

The Pfu RPP29 construct contains 127 residues with the first 35 residues unique to the Pyrococcus family (Fig. 4.1). In the absence of RPP21, the first 47 residues of Pfu RPP29 do not give signals in the 2D NMR spectrum. This indicates that this segment may not be required for maintaining the structure of Pfu RPP29 \textit{per se}. A limited trypsin proteolysis experiment showed that a trypsin-resistant core was not sensitive to digestion after 40 min, while the remaining parts can be quickly removed within 10 min. ESI-MS result revealed a tryptic fragment (Fig. 4.2), which is composed of residues 42-120. It is noteworthy that trypsin only cleaves after Arg and Lys residues, and the next Arg/Lys residue after K41 is K52. Based on the 2D spectrum, K52 is in the region which is well folded. Therefore, the unstructured N-terminus might be beyond K41, but can not
be cleaved by trypsin, consistent with the NMR observations.

**Pfu RPP29 derivatives**

The crystal structure of *Pho* RPP29 was solved with an N-terminal deletion mutant: RPP29Δ42 and RPP29Δ36, since crystals could not be obtained with the wild type protein.[48] Based on these observations, two N-terminal deletion mutants of *Pfu* RPP29 were engineered, RPP29Δ42 and RPP29Δ36, in which the first 42 and 36 residues are removed, respectively. 2D NMR spectra confirmed that the structured core is conserved in both mutants (data not shown), consistent with the solution structures of *Mth*[47] and *Afu* RPP29.[49] However, the addition of the unlabeled RPP21 did not induce significant spectral changes, indicating the deletions eliminated important binding determinants (data not shown). Thus, the interaction with RPP21 requires the N-terminal region of RPP29, contrary to the prediction based on the RPP29 sequence alignment.

Given the goal to generate a smaller construct of RPP29 for NMR data collection and subsequent structure determination of the RPP29-RPP21 complex, we made another five mutants with different length deletions at the N-terminus, namely RPP29Δ5, RPP29Δ13, RPP29Δ17, RPP29Δ24 and RPP29Δ31 (Fig. 4.1). RPP29Δ31 corresponds to the construct used in the crystallization of *Pho* RPP29.[48] This series of RPP29 derivatives also allowed us to more precisely explore the function of the N-terminus of RPP29. The ability of each mutant to bind RPP21 was examined by NMR spectroscopy and ITC, and its effect on RNase P activity was tested by *in vitro* reconstitution assay.

**NMR assays**
As observed for RPP29Δ36 and RPP29Δ42, the other five RPP29 deletion mutants also adopt the same hydrophobic core, indicated by nearly identical 2D NMR spectra (Fig. 4.2a-c, data shown for RPP29Δ17 and RPP29Δ24 only). Among them, only three mutants resulted in dramatic spectral changes in the presence of unlabeled RPP21: RPP29Δ5, RPP29Δ13 and RPP29Δ17 (Fig. 4.2e, data not shown for RPP29Δ5 and RPP29Δ13). A handful peaks in RPP29Δ24's bound spectrum have very small chemical shift perturbations, but the rest of the peaks remain at the same position between free and bound spectra of RPP29Δ24. This suggests weak binding between RPP29Δ24 and RPP21 (Fig. 4.2f). The ability to bind RPP21 is completely lost in RPP29Δ31 (data not shown), as for RPP29Δ36 and RPP29Δ42. The reverse titration (unlabeled RPP29 constructs titrated into U-[15N]-RPP21) confirmed the same result (data not shown).

**ITC**

In order to assess the RPP29 derivatives more quantitatively and inform subsequent NMR studies, we measured the binding affinity between RPP29 and RPP21 using ITC. Based on the initial trials with RPP21V14 and RPP29 derivatives (data not shown), careful characterizations were carried out on RPP21WT and RPP29WT, RPP29Δ17, and RPP29Δ24. Consistent with NMR data, RPP29Δ17 resembles RPP29WT, with comparable stoichiometry (N) and ΔH (enthalpy change), though it binds to RPP21 a little weaker (association constant (K_a) of 0.352 μM for RPP29Δ17 and 0.250 μM for RPP29WT) (Fig. 3.2 and Table 3.1). On the other hand, ITC was not able to detect binding between RPP29Δ24 and RPP21 under the same conditions. Therefore, we concluded that RPP29Δ17 would be the best construct for structure determination of RPP21-RPP29 complex. As illustrated in Chapter 3, this construct almost completely prevents the otherwise observed signal loss in RPP29Δ17 spectra with unlabeled
RPP21V14, allowing us to determine the structure of the RPP29Δ17-RPP21V14 complex. As observed in the wild type RPP29-RPP21 complex, residues 1-17 in RPP29 do not adopt a stable conformation, confirming our rationale of choosing RPP29Δ17 for structure determination of RPP21-RPP29 complex. The structure of the complex also provides a structural explanation for the large affinity difference for RPP21 binding to RPP29Δ17 and RPP29Δ24 (Fig. 3.8b).

Reconstitution assays

To complete our understanding of the RPP29 derivatives, steady state kinetic assays using in vitro reconstituted enzyme were employed to examine the effect of RPP29 N-terminal deletions on RNase P activity in both partially (RPR+RPP21+RPP29 derivative) and fully (RPR+RPP30+POP5+ RPP21+RPP29 derivatives) reconstituted enzyme. Here, the reconstituted RNase P using RPP29WT or derivatives will be denoted as RPP29WT RNase P or RPP29ΔX RNase P. When all the components are present (fully reconstituted Pfu RNase P), RPP29Δ17 RNase P and RPP29Δ24 RNase P has 83% and 54% of the RPP29WT RNase P activity, while RPP29Δ31 RNase P has less than 1% of the catalytic activity of the wild type enzyme (Fig. 4.4 and Table 4.2). This is expected because the deletion of the N-terminal 31 residues eliminates its ability to bind RPP21. Thus, the formation of the RPP21-RPP29 complex must be crucial for the RNase P activity. However, recall that RPP29Δ24 is also severely impaired for binding to RPP21, the 54% of activity retained by RPP29Δ24 RNase P suggests that the RPR may rescue the defect of RPP29Δ24 in RPP21 binding. As mentioned in Chapter 2, one putative RNA binding interface contains elements of both proteins, and proposed to interact with the S-domain of RPR. In the absence of the RPR, the N-terminus (residues 25-35) of RPP29Δ24 can not adopt a stable conformation for RPP21 binding; while in the presence of RPR, the
large RNA may serve as a platform to recruit both proteins and then induce the appropriate conformation of each protein required for interaction.

Strikingly, when the activity was measured on the partially reconstituted RNase P (RPR + RPP21 + RPP29 derivatives), RPP29Δ17RNase P possesses only 41% of the activity, while RPP29Δ24RNase P has 84% of the activity (Table 4.2). However, it is noteworthy that the RNase P activity is severely impaired in the absence of RPP30-POP5 complex. The turnover rate of the partially reconstituted RNase P (RPP + RPP21 + RPP29, 0.188 min⁻¹) is much slower than the fully reconstituted RNase P (RPR + 4 RPPs, 9.768 min⁻¹). Previous NMR assays have shown no detectable interaction between two protein pairs in the absence of the RPR (data not shown) at 0.3 mM (protein). The footprinting results also indicate that the RPP21-RPP29 and the RPP30-POP5 complexes exclusively interact with the S- and the C-domains of RPR, respectively. Thus, in the partially reconstituted RNase P, the interaction between the S-domain and the RPP21-RPP29 complex seems to work better for RPP29Δ24 than for RPP29Δ17. But when the RPP30-POP5 complex is present, the C-domain must be better folded, and somehow has a positive influence on the better folding of the S-domain as well, which results in a better enzymatic activity of RPP29Δ17RNase P than that of RPP29Δ24RNase P. This might suggest direct RNA-protein interactions or indirect protein-protein interactions between two binary complexes via RNA. Further experiments are required to resolve these observation (unpublished data by I-Ming Cho, Gopalan lab).

CONCLUSION
Though described after the structure determinations of *Pfu* RPP29-RPP21 (Chapter 2) and RPP29Δ17-RPP21V14 (Chapter 3) complexes, the identification of the minimal domain of *Pfu* RPP29 for assembly with *Pfu* RPP21 (RPP21V14) was an early success during this project. Without it, no NMR spectra would have been able to recorded on RPP29Δ17-RPP21V14 complex, and thus no structure of RPP29Δ17-RPP21V14 would have been solved to facilitate the structure determination of RPP29-RPP21 wild type complex. This set of experiments also serves as a good example of thorough examination of the protein mutants from structural, thermodynamic, and functional perspectives, using NMR spectroscopy, ITC, and kinetic assays, respectively.

Wen-Yi Chen in the Gopalan lab has discovered that a functional defect resulting from a mutation in the RPR P4 region can be recovered by the addition of the POP5-RPP30 complex (unpublished data, Wen-Yi Chen, Gopalan lab). Conversely, a defect in the RPPs can also be rescued by the RPR, as demonstrated by the RPP29Δ24\(^{\text{RNase P}}\) results. The removal of the first 24 residues largely abolishes its ability to bind RPP21, and yet in the *in vitro* reconstitution assay, the enzyme reconstituted from RPP29Δ24 retains 54% or 84% of the catalytic activity of partially- and fully-reconstituted RNase P, suggesting that the interaction between RPP29Δ24 and RPP21 is recovered by the presence of the RPR. Thus, the RPR is not only responsible for the enzymatic catalysis, but also serves as a structural platform for recruiting the RPPs to form the functional RNase P holoenzyme. These observations of mutual rescue between RPR and RPPs has revealed the interdependent relationship among all components in archaeal RNase P, improving our understanding of the architecture of the RNase P holoenzyme.
Figure 4.1. Sequence alignment of selected RPP29 homologs from Archaea and Eukarya. The alignment was generated with CLUSTALW, and illustrated using ESPRIT2.2, in which red letters indicate a global similarity score of 0.7, and red boxed letters indicate invariant residues. Secondary structural elements represented in cartoon are observed in the crystal structures of Pfo RPP29 in the complex. Aligned sequences are from Pyrococcus furiosus (NCBI entry NP_579545), Pyrococcus horikoshii (NP_143607), Methanobacterium thermoautotrophicum (10QK_A), Methanococcus jannaschii (NP_247439), Archaeoglobus fulgidus (1TSF_A), Saccharomyces cerevisiae (NP_009816), Schizosaccharomyces pombe (NP_588479) and Homo sapiens (NP_006618). Black arrows show the N-terminal deletion mutants of Pfu RPP29 we constructed. The light red arrow shows the Pho N-terminal deletion mutant Pho RPP29N31 used for crystallization, the electron density map of which starts at residue 36, indicated by the dark red arrow. The blue arrow illustrates where the conserved core starts, identified by limited trypsin proteolysis.
Figure 4.2. Mass spectrometric analysis of trypsin-digested Pfu RPP29. ESI-MS data analysis was performed by Ross Wilson. Upon injecting into the ESI Mass Spectrometer, each peptide is multiply protonated ions with different mass to charge (m/z) ratios. ESI-MS results on the tryptic fragments of Pfu RPP29 illustrated the hydrophobic core is composed of residues 42-120. Some of the observed ions could be explained by partial oxidation of the protein.
Figure 4.3. NMR binding assay of *Pfu* RPP29 N-terminal deletion derivatives. Both RPP29Δ17 (c) and RPP29Δ24 (e) adopt the same structured core as the wild type (a), indicated by the almost identical 2D spectra of the free proteins. Titration of RPP21 into RPP29 only induces dramatic spectral changes in the spectra of RPP29WT (b) and RPP29Δ17, but not in the spectrum of RPP29Δ24 (f).
Figure 4.4. *In vitro* reconstitution assay of the fully reconstituted *Pfu* RNase P. The experiments were performed under the optimal condition for the fully reconstituted *Pfu* RNase P (10 nM). The final concentration of the substrate (ptRNA\(^\text{Tyr}\)) was kept at 625 nM. (a) Samples were taken at 0.5, 1, 1.5, 2 and 2.5 min for RPP29WT and RPP29Δ17, at 1, 2, 3, 4 and 5 min for RPP29Δ24, at 10, 20, 30, 40 and 50 min for RPP29Δ31, and were run on a denaturing gel. (b) Initial velocities were obtained from linear fits of product concentration vs. time, converted to turnover number, and were reported as “relative activity” towards the wild type RNase P activity (Table 4.2). Though RPP29Δ24 has been shown no detectable interaction with RPP21 by NMR and ITC, it still retains 54% of the RNase P activity in the reconstitution assay, suggesting a rescue from the RPR to restore RPP21 binding (unpublished data, I-Ming Cho, Gopalan lab).
Table 4.1. Oligonucleotide primers used for cloning the Pfu RPP29ΔX derivatives.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PfuP29Δ5-F</td>
<td>5’-cctagtccATGGAAGAACGTGAGAACTTCAG-3’</td>
</tr>
<tr>
<td>PfuP29Δ13-F</td>
<td>5’-gcatttccATGGGAGATCACACAGGGATCG-3’</td>
</tr>
<tr>
<td>PfuP29Δ17-F</td>
<td>5’-cctcgaccATGGGATCGTATCAAGAAATTATTG-3’</td>
</tr>
<tr>
<td>PfuP29Δ24-F</td>
<td>5’-gcatttccATGGGTGAGAATGAGAATAGAACTTCAG-3’</td>
</tr>
<tr>
<td>PfuP29Δ31-F</td>
<td>5’-cctaaaccATGGGAGCTCAGAGGTAGAGTAAAC-3’</td>
</tr>
<tr>
<td>PfuP29Δ36-F</td>
<td>5’-cgtggaaccATGGGTAGAGTAAACATAGATATGGAACCGAAGTTTATTG-3’</td>
</tr>
<tr>
<td>PfuP29Δ42-F</td>
<td>5’-cgaatttccATGGGATCGTATCAAGAATAGAACTTCAG-3’</td>
</tr>
<tr>
<td>PfuP29C127-R</td>
<td>5’-cctaaactcgagTCAAGGCAAGCTAGAGTTTTTACGCCAACGC-3’</td>
</tr>
</tbody>
</table>

Start and stop codons are highlighted in bold, and BamHI (in Forward primers, denoted as F) and XhoI (in Reverse primer, denoted as R) restriction sites are underlined.

Table 4.2. Relative activity of partially and fully reconstituted RNase P (unpublished data, I-Ming Cho, Gopalan lab).

<table>
<thead>
<tr>
<th>RPP29 derivative</th>
<th>Fully reconstituted RNase P (RPR+RPP30+POP5+RPP21+RPP29 derivative)</th>
<th>Partially reconstituted RNase P (RPR+RPP21+RPP29 derivative)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type (WT)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>RPP29Δ17</td>
<td>0.83</td>
<td>0.41</td>
</tr>
<tr>
<td>RPP29Δ24</td>
<td>0.54</td>
<td>0.84</td>
</tr>
<tr>
<td>RPP29Δ31</td>
<td>0.02</td>
<td>0.12</td>
</tr>
</tbody>
</table>

* The relative activities were calculated from three independent experiments with standard deviations less than 20%.
CHAPTER 5

DISSECTING BINDING-COUPLED FOLDING IN THE Pfu RPP21-RPP29 INTERACTION BY ISOTHERMAL TITRATION CALORIMETRY

INTRODUCTION

Macromolecular recognition is an essential process in biological functions, such as enzymatic catalysis, gene regulation, and signal transduction. Though high resolution structures can reveal the details of specific noncovalent interactions that stabilize the structure of macromolecular assemblies, the physical principles governing such interactions remain poorly understood.\[112\] For a binding reaction:

\[ M + L \leftrightarrow ML \]

with equilibrium association constant \( K_A \):

\[ K_A = \frac{[ML]}{[M][L]} \]

a spontaneous complex formation requires a favorable binding free energy, \( \Delta G < 0 \), driving the transition from the free components to the final complex. The total \( \Delta G \) also reflects the binding affinity (\( K_A \)) according to:

\[ \Delta G = -RTlnK_A \] \hspace{1cm} (1)

\( \Delta G \) can be dissected into fundamental thermodynamic components: binding enthalpy (\( \Delta H \)), binding entropy (\( \Delta S \)), based on the Gibbs equation:

\[ \Delta G = \Delta H - T\Delta S \] \hspace{1cm} (2)
Favorable (negative) contributions to $\Delta H$ of binding arise from forming new interactions such as hydrogen bonds, van der Waals contacts and favorable electrostatic pairings. $\Delta S$ is a measure of change in system order upon binding, reflecting a balance between the favorable release into bulk solvent of ordered surface water molecules and unfavorable loss of translational, rotational and vibrational degrees of freedom.[113; 114] Importantly, this last term includes losses of conformational flexibility in both binding partners.

However, Equation 2 does not illustrate the observation that $\Delta G$ is relatively unchanged over a broad range of temperatures. This is because the enthalpy and entropy terms are dependent on temperature, as described by the heat capacity change of the interaction ($\Delta C_p$).

$$\Delta H(T) = \Delta C_p(T - T_n) \quad (3)$$
$$\Delta S(T) = \Delta C_p \ln(T/T_s) \quad (4)$$

where $T_n$ and $T_s$ are the temperatures at which the binding enthalpy and entropy are zero, respectively. As the most information-rich thermodynamic term, $\Delta C_p$ contains information including the hydration of hydrophobic groups, hydrogen bonding, electrostatics, protein conformational entropy, vibrational terms, and changes in equilibrium.[115] A large negative $\Delta C_p$ of binding has been shown to be diagnostic of binding-coupled folding and has been used to estimate the number of residues that fold upon binding.[85] However, $\Delta C_p$ also contains contributions from linked equilibria other than binding-coupled folding, such as ion and proton linkage, which need to be accounted for in order to quantify coupled folding.
Isothermal titration calorimetry (ITC) can directly measure the heat associated with complex formation (i.e., ΔH), and allows one to deduce ΔG, ΔS and ΔC_p, so as to offer a full understanding of the energetics of complex formation. A brief description of the technique is given for a 1:1 binding reaction between two molecules L and M.[116] The titration experiment is performed by injecting a small amount of a solution containing L in the syringe into the sample cell containing a solution with M. Upon interaction, heat may be released (exothermic) or absorbed (endothermic), leading to a temperature change in the sample cell. The ITC instrument continuously monitors the temperature difference (ΔT) between the sample cell and a reference cell (usually containing water), and changes the electric power supplied to the sample cell in order to maintain a fixed ΔT. Thus, each injection results in a peak-shaped instrumental response (μcal s⁻¹) as heat is released / absorbed and the instrument compensates for the temperature difference. The next injection starts after the equilibrium is re-established. At the beginning of titration, [L] << [M], the injected L will bind to M and release / absorb heat in proportion to the amount of binding. As the binding sites on M become saturated upon subsequent addition of L, reducing fractions of the injected L will bind, releasing /absorbing less heat. Once M is saturated, there is no further heat due to binding, and only minimal heat will arise from dilution and mechanical effects. Fig. 5.1 shows typical isotherms for exothermic (a) or endothermic (b) reactions. The area of each peak (q_i) is integrated, adjusted for the heat of dilution, and converted to molar enthalpy (ΔH, in kcal mol⁻¹) using the total concentration of M and L at the ith injection and the cell volume (v). The resulting values are plotted as a function of the molar ratio and fit to a one binding site model. The binding enthalpy (ΔH) and binding association constant K_A can be obtained by fitting the heat at each injection to Equation 4 :[113]
where [L] is the concentration of free molecule L and K_A is the binding constant. Once ΔH and K_A are obtained, ΔG and ΔS can be deduced based on Equations 1 and 2, respectively. Therefore, a single titration experiment can yield most of the thermodynamic parameters (ΔG, ΔH and ΔS) and K_A. By repeating the experiment at different temperatures, ΔC_p can be determined as well according to Equation 3.

After determining the solution structure of the Pfu RPP21-RPP29 complex, we used the program Structure Thermodynamics Calculation (STC) to calculate the change in accessible surface area (ASA) and the energetics of the RPP21-RPP29 complex formation.[95] The total change in ASA is distinguished between polar ASA (ASA_p) and nonpolar ASA (ASA_np), which make unfavorable and favorable contributions to ΔC_p, respectively. Based on the solution structure, there is a burial of approximately 2400 Å² upon interaction, corresponding to a ΔC_p of -615.4 cal mol⁻¹ K⁻¹. Similar results are found in the Pho RPP21-RPP29 complex, with a burial of 2430 Å² of ASA,[97] and a ΔC_p of -503.918 cal mol⁻¹ K⁻¹ (calculated using STC). However, the ΔC_p reported by STC is solely from the ASA difference between the free and bound states, and does not take binding-coupled folding into consideration, which also contributes to ΔC_p. From NMR studies, we observed a large portion of the Pfu RPP21-RPP29 complex only becomes ordered upon interaction, including residues 17-48 and 116-123 of RPP29 and residues 9-18 of RPP21.

To understand the thermodynamic behavior of the RPP21-RPP29 interaction, we performed a series of ITC experiments to dissect the binding energetics in terms of ΔG, ΔH, ΔS and ΔC_p. We discovered that the ΔC_p is almost twice (-1138.65 kcal mol⁻¹) as
large as the prediction from the structure, suggesting other linked equilibria, such as binding-induced folding, or ion and proton linkage effects are coupled to the protein-protein interaction. Among these linked equilibria, we found that ion linkage does not contribute to $\Delta C_p$ significantly, while two protons are released during complex formation. Based on the experimental pH, we proposed that the only two His residues in the binding interface are deprotonated upon complex formation, in order to take part in hydrogen bonding. After correcting for linked protonation, the measured $\Delta C_p$ corresponds to fifty residues being stabilized by the protein-protein interaction; this number is in a very good agreement with our NMR observations. These analyses allowed us to correlate the structure and thermodynamics of the RPP21-RPP29 interaction, revealing not only how but also why the interaction takes place.

MATERIALS AND METHODS

Protein preparation

The overexpression and purification of *Pfu* RPP21 and RPP29 followed the same procedures described in Chapter 2. Both proteins (<2 mL each) were transferred into a membrane tubing with 3,500 Da cutoff, and dialyzed in the same beaker containing 0.5-0.7 L of ITC buffer for at least 6 hours to avoid buffer mismatch during ITC. The dialysis was repeated again for at least 6 hours to guarantee the complete exchange into the desired buffer conditions. The protein concentrations were estimated using the extinction coefficient, 37,470 M$^{-1}$cm$^{-1}$ (ProtParam: http://ca.expasy.org/tools/protparam.html) at 280 nm for *Pfu* RPP29, and 16,180 M$^{-1}$cm$^{-1}$ (ProtParam) at 280 nm for *Pfu* RPP21.

Isothermal titration calorimetry
All ITC experiments were performed on a MicroCal VP-ITC instrument, with the cell containing one protein at a concentration of [M], and the syringe containing the other protein at a concentration of [L]. In order to obtain accurate $K_A$ from ITC, the Wiseman c value $\equiv K_A \times [M]$ should be kept between 10 and 100.[112] Preliminary experiments have shown that the binding association constant $K_A$ of the RPP21-RPP29 interaction is $4 \times 10^6$ M$^{-1}$ at 55°C (Chapter 3). Therefore, the proteins in the cell and in the syringe were prepared at concentrations of ~20 μM and ~200 μM respectively, resulting in a Wiseman c-value of 80. The samples were thoroughly degassed (5 min, room temperature) prior to titration. Each experiment consisted of a first injection of 3 μL and following injections of 5-6 μL with a 400 s interval between each injection.

The results were analyzed using the Origin (V.7 SR4) software package provided by MicroCal. The heat of dilution was obtained by averaging the last few injections after saturation, and that value was removed from the integrated heat pulse from each injection. By fitting the adjusted heat pulses to a one-site binding model (Equation 4) in a nonlinear least squares manner, Origin determines the stoichiometry ($n$), binding constant ($K_A$) and enthalpy ($\Delta H$) based on the given concentrations of each protein and the cell volume. Then, $\Delta G$ and $\Delta S$ were deduced from Equation (1) and (2), respectively. The reported errors correspond to the deviation of the nonlinear least squares fit to the data points on the titration curve.

**Measurement of $\Delta C_p$ by ITC**

Both RPP21 and RPP29 were dialyzed twice against the standard ITC buffer (20 mM Cacodylate (sodium salt), pH 6.7, 10 mM KCl, 0.3 mM ZnCl$_2$, and 0.02% NaN$_3$, ionic strength I= 67.6 mM). Cacodylate was chosen due to its small ionization enthalpy (-
0.717 kcal mol\(^{-1}\)). This results in low temperature sensitivity of this buffer, 0.04 pH unit over a temperature range from 10°C to 45°C. Titrations of RPP29 (200 μM) into RPP21 (20 μM) were performed in this temperatures range, with an interval of 5°C. The experiment at 25°C was omitted due to a \(\Delta H\) close to zero. If the temperature dependence of \(\Delta H\) can be fit to a straight line, the slope is the \(\Delta C_p\) of the interaction according to Equation 3. Curvature in this plot implies a temperature dependent \(\Delta C_p\) for the system. It is noteworthy that the measured \(\Delta C_p\) is a global value, which can include all of the contributions from 1) binding-coupled folding, 2) ion linkage and 3) linked protonation.\[118\]

**Ion linkage**

If the complex formation between molecules M and L involves release of ions (I), the binding equilibrium can be expressed as:

\[
M \cdot n_1 I_{\text{anion}} + L \cdot n_2 I_{\text{cation}} \leftrightarrow ML + n_1 I_{\text{anion}} + n_2 I_{\text{cation}}
\]

where \(n_1\) and \(n_2\) are the numbers of released anions and cations, respectively. The equilibrium association constant \((K_A)\) can then be defined by:

\[
K_A = \frac{[ML]}{[M][L]} \left[ I_{\text{anion}} \right]^{n_1} \left[ I_{\text{cation}} \right]^{n_2}
\]

\[
\log K_A = \log \left( \frac{[ML]}{[M][L]} \left[ I_{\text{anion}} \right]^{n_1} \left[ I_{\text{cation}} \right]^{n_2} \right) = \log K_A^{\text{complex}} + n_1 \log [I_{\text{anion}}] + n_2 \log [I_{\text{cation}}]
\]

where \(K_A^{\text{complex}}\) is the equilibrium association constant only describing the complex formation. Assume that the concentrations of anions and cations are the same, the equation above can be simplified into:

\[
K_A = \log [K_A^{\text{complex}}] + (n_1 + n_2) \log [I]
\]  \(5\)
where $n_1 + n_2$ reports the number of ion linkage coupled with protein-protein interactions. Ion linkage can be identified by carrying out ITC titrations at different salt concentrations. If a salt bridge is formed or broken upon complex formation, the increase in salt concentration will impair the $K_A$ by inhibiting electrostatic interactions. Thus, by plotting $\log K_A$ vs. $\log[\text{salt}]$, the numbers of ion released or adsorbed upon formation of a complex can be estimated from the slope of the fit line. On the other hand, if $K_A$ is insensitive to salt concentration, this indicates no ion transfer between the proteins and bulk solvent upon binding.

We performed the experiments in the standard ITC buffer with four different KCl concentrations: 10 mM ($I = 67.8$ mM), 50 mM ($I = 147.8$ mM), 100 mM ($I = 247.8$ mM), 150 mM ($I = 347.8$ mM) and 200 mM ($I = 447.8$ mM). When dialyzing 200 μM RPP29 into the buffer containing 150 mM KCl, we observed some precipitation, resulting in only 128 μM of RPP29 remaining in solution. However, RPP21 appears to be soluble in the buffer containing 150 mM KCl at concentration of 200 μM. Thus, titrations were performed in reverse by injection of RPP21 (150-200 μM) into RPP29 (15-20 μM). No precipitation was observed prior to or after titration. The same sets of experiments were carried out at 55°C and 10°C, respectively, in order to investigate the temperature dependence of this ion linkage effect. Furthermore, we performed the ITC experiments at 150 mM KCl over a wide range of temperatures (10°C to 45°C), and determined the $\Delta C_p$ at this second KCl concentration, in order to correct $\Delta C_p$ for linked ion binding.

**Proton linkage**
The contribution from protonation / deprotonation to $\Delta C_p$ has been long discovered and well documented. Some reviews have dedicated to this particular equilibrium coupled with protein-protein interactions.[112; 119; 120] In Baker and Murphy’s paper, [119] such binding-coupled proton linkage is illustrated by a 1:1 complex (ML) binding a single proton:

\[
\begin{align*}
M & \leftrightarrow K_{p-p} ML \\
K_p^t & \uparrow \downarrow K_p^c \\
M^+ & \leftrightarrow ML^+
\end{align*}
\]

where the protonated M and ML are represented as $M^+$ and $ML^+$. $K_{p-p}$ is the binding constant for the deprotonated protein interaction. The protonation of M and ML are described by the proton binding constant $K_p^t$ and $K_p^c$, respectively, which are equal to $10^{pK_a}$ of the ionizing group. When more than one proton is associated with protein binding, the observed protein binding constant $K_{obs}$ can be expressed as:

\[
K_{obs} = K_{p-p} \left( \frac{1 + K_p^c [H^+]}{1 + K_p^t [H^+]} \right)
\]

This indicates that proton linkage is the change in the proton affinities of the protein between free and bound states. Moreover, the number of protons released or absorbed upon binding $N_{H^+}$ can be described by the fractional saturation of protons in the free state ($H^f$) and the bound state ($H^c$):

\[
N_{H^+} = H^c - H^f = \frac{K_p^c \alpha_{H^+}}{1 + K_p^c [H^+]} - \frac{K_p^t [H^+]}{1 + K_p^t [H^+]}
\]

In this case, the measured binding enthalpy ($\Delta H_{obs}$) from ITC experiments is composed of both $\Delta H$ from protein-protein interaction and protonation / deprotonation:
\[ \Delta H_{\text{obs}} = \Delta H_{\text{bind}} + N_{\text{H}^+} \Delta H_i \] (6)

where \(\Delta H_{\text{bind}}\) corresponds to the enthalpy change purely from protein binding and \(\Delta H_i\) is the ionization enthalpy of the buffer. Therefore, by performing ITC experiments in buffers with different \(\Delta H_i\) (at a given pH), the slope of \(\Delta H_{\text{obs}}\) vs. \(\Delta H_i\) yields the number of protons transferred between the proteins and the solvent. The sign of the slope is an indicator of proton uptake (slope > 0) or release (slope < 0).[112]

Once the \(N_{\text{H}^+}\) is determined, the corrected \(\Delta H_{\text{obs}}\) (\(\Delta H_{\text{bind}}\)) reports a \(\Delta C_p\) that does not contain proton linkage effect. When correcting the \(\Delta H_{\text{obs}}\) according to Equation 6, we also adjusted the \(\Delta H_i\) based on the buffer \(\Delta H_i^{\text{ref}}\) and \(\Delta C_p\), because \(\Delta H_i\) is also temperature dependent.[121] It is noteworthy that \(\Delta H_{\text{bind}}\) still contains both the intrinsic enthalpy change for protein binding (\(\Delta H_{\text{intr}}\)) and the enthalpy change for the protonation of the inoizable group(s) on the proteins (\(\Delta H_{\text{prot}}\)).

The titrations of RPP21 (~150 μM) into RPP29 (~15 μM) were carried out in three buffers, cacodylate (\(\Delta H_{\text{ion}} = -0.717\) kcal mol\(^{-1}\)), MOPS (\(\Delta H_{\text{ion}} = 5.043\) kcal mol\(^{-1}\)) and ACES (\(\Delta H_{\text{ion}} = 7.273\) kcal mol\(^{-1}\)) with 10 mM KCl, 0.3 mM ZnCl\(_2\) and 0.02% NaN\(_3\) (I = 67.8 mM for all three buffers). The pH was carefully adjusted to 6.7 for all three buffers at 55°C.[117] Uncertainty in measured value were obtained by a nonlinear least square fit to Equation 4.

Tables of thermodynamic parameters obtained from ITC experiments

The thermodynamic parameters from each titration, \(\Delta H\), \(K_A\), and the deduced \(\Delta G\) and \(\Delta S\), are tabulated in Appendix E, along with the standard deviations from the nonlinear square fit of the data to the titration curve (Origin).
RESULTS AND DISCUSSION

Enthalpy-entropy compensation

The RPP21-RPP29 interaction is endothermic below 25°C and exothermic above 25°C, reflecting a negative $\Delta C_p$. At about 25°C, the binding enthalpy approaches zero (Fig. 5.2). Likewise, the $\Delta S$ of the interaction is unfavorable ($\Delta S < 0$) below 35°C and favorable ($\Delta S > 0$) above 35°C (Fig. 5.2). Because temperature changes have nearly equal but opposite effects on $\Delta H$ and $T\Delta S$, the compensation between enthalpy and entropy results in relative insensitivity of $\Delta G$ over the experimentally sampled temperatures, with an optimal $\Delta G$, corresponding to tightest binding, at about 30°C. The linear fit of $\Delta H$ vs. temperatures reports a $T_h$ of 25.36°C (298.51 K), consistent with our observation of negligible $\Delta H$ of RPP21-RPP29 titration at 25°C.

Temperature-independence of $\Delta C_p$

On the assumption of a rigid body interaction, calculations based on the solution structure of the Pfu RPP21-RPP29 complex using STC found a $\Delta C_p$ of -615.4 cal mol$^{-1}$K$^{-1}$ from burial of approximately 2400 Å$^2$ of solvent accessible surface area (ASA) during complex formation. In contrast, we obtained a $\Delta C_p$ of -1133.48 cal mol$^{-1}$ K$^{-1}$ from the ITC experiments carried out from 10°C to 45°C (Fig. 5.3). This value is almost twice the prediction from STC, suggesting that other linked equilibria are coupled with the interaction. In heteronuclear NMR studies, we observed that fifty residues disordered in the free proteins become structured upon complex formation: residues 17-48 and 116-123 of RPP29 and residues 9-18 of RPP21. As a large negative $\Delta C_p$ can be a thermodynamic signature of binding-coupled folding,[85] our measured $\Delta C_p$ appears to reflect the induced fit occurring in the RPP21-RPP29 interaction. Since $\Delta H$ was found to
change linearly with temperature in the 10-45°C range (Fig. 5.3), this indicates that ΔC_p is temperature independent over that range. It is noteworthy that the measured ΔC_p may contain contributions from other linked equilibria, including possible ionization and protonation.[85]

**Ion linkage**

The binding interface of the *Pfu* RPP21-RPP29 complex is characterized mostly by hydrophobic interactions surrounding prominent intermolecular electrostatic contacts (Fig. 5.4). In particular, two intermolecular salt bridges are proposed to form upon binding between E47^{RPP29} and R17^{RPP21} and between D72^{RPP29} and R38^{RPP21}. Based on the high conservations of these four residues in the archaeal RPP29 and RPP21, these polar interactions may play an important role in interface formation. Though due to the limited NOEs, the side chains of these residues could not be constrained to adopt stable conformations, some structures in the ensemble are within the distance to allow salt bridges (Fig. 5.4b). In the crystal structure of the *Pho* RPP21-RPP29 complex, the distances between the heteronuclei of E47^{RPP29}-R17^{RPP21} (R22 in *Pho* RPP21) and D72^{RPP29}-R38^{RPP21} (R43 in *Pho* RPP21) were 2.7 Å and 2.6 Å, respectively. If these residues are bound to counter ions in the free proteins, we expected high salt concentration to compete with intermolecular ion pairing, and impair the binding. Contrary to that prediction, the RPP21-RPP29 interaction becomes tighter at higher salt concentration, with a 15-fold increase of the binding affinity in 150 mM KCl compared to 10 mM KCl, at 55°C. A linear fit of log K_A vs log[KCl] yielded a slope of 1.4, indicating that uptake of 1.4 ions is linked to binding (Fig. 5.5).
The enhanced binding affinity at high salt concentrations are observed during protein-DNA interactions in the halophilic organisms[122], which use ions to shield the polyelectrolyte nature of DNA, and to mediate DNA-protein interactions. Both RPP21 and RPP29 are highly positively charged at neutral pH, with a pI over 10. The surface of the entire complex is largely covered with basic residues, and one putative RNA binding interface (site 1) is constituted by both proteins (Fig. 5.4c). Therefore, a cluster of basic residues might interact in an intermolecular manner, or reorient in the presence of the other protein, resulting in charge repulsions. Ions then might bind to these surface residues to allow the packing of much electropositivity. However, the ion binding is not site-specific, and the bound ions are constantly exchanging with the ions in the solvent, resulting in an non-integral number of absorbed ions.

Alternatively, ions may help to fold more disordered region in the free protein by preventing unfavorable charge repulsion. The electronegativity on RPP29 interface is contributed by E47 and D72, while the electropositivity on RPP21 interface is from R17 and R38. However, both E47RPP29 and R17RPP21 are disordered in the free protein, probably due to the unfavorable packing of same charges. At higher salt concentration, the polyelectrolytic effect may be prevented by recruiting counter ions bound to the charged groups. Ion binding allows more content of protein to be able to fold in the free state, resulting in tighter binding to the other protein. It is noteworthy that the bound ions must be released when the salt bridge forms, thus the measured ion linkage is a combination of the polyelectrolytic effect (slope > 0) and the electrostatic contacts.

The ITC experiments in 10-150 mM KCl were repeated at 10°C, and the fitted data report a similar salt dependence on binding (Fig. 5.5). Interestingly, even higher affinity was observed at 10°C, 3-5 fold compared to the $K_a$ from 55°C at same salt
concentrations. One possibility is that though both proteins contain unstructured termini in their free states, the ordered cores may become partially unfolded at higher temperature, leading to weaker interaction. Considering that Pfu lives in 95°C, it is quite intriguing that the proteins appear to become less structured at higher temperature (55°C). The biological implication of this observation is still unclear, but it may reflect differences between the cellular conditions in the archaeon, and those used in our in vitro experiments.

To investigate the contribution of ion linkage to the ΔC<sub>p</sub>, we performed the titrations in 150 mM KCl at temperatures ranging from 10°C to 55°C (with an interval of 10°C). The ΔC<sub>p</sub> obtained under these conditions (ΔC<sub>p</sub> = -1127.11 ± 231.1 kcal mol<sup>-1</sup>K<sup>-1</sup> in 150 mM KCl) is similar to that obtained from condition of 10 mM KCl (ΔC<sub>p</sub> = -1092.32 ± 161.4 kcal mol<sup>-1</sup>K<sup>-1</sup>), within experimental error. This indicates that this linked ionization does not contribute to ΔC<sub>p</sub> significantly (Fig. 5.6).

It is noteworthy that the salt dependence of the K<sub>a</sub> probably reflects opposing effects from hydrophobic interactions and intermolecular ion pairs. However, currently we do not have the data to allow us to dissect them.

**Proton linkage**

Similarly, the proton linkage was characterized by performing ITC experiments in buffers of cacodylate, MOPS and ACES at pH 6.7. When plotting ΔH vs. ΔH<sub>i</sub>, a linear fit renders a slope of about -2 (slope = -2.1640 ± 0.1314, Fig. 5.7), indicating the transfer of two protons from the proteins to the solvent upon binding. After correcting the measured ΔH (in cacodylate) from this linked deprotonation according to Equation 6, we obtained a corrected ΔC<sub>p</sub> of -1.09232 kcal mol<sup>-1</sup> K<sup>-1</sup> (Fig. 5.3). The small correction is due to the small ΔH<sub>i</sub> of cacodylate buffer (0.04 pH unit change within 10-45°C)
At pH 6.7, the two deprotonated groups can most likely be attributed to His imidazole side chains ($pK_a = 6$).[123] There are two His residues buried in the binding interface of RPP21-RPP29 complex, H34 and H46, both from RPP29. Other His residues such as H60 in RPP29, H67, H87 and H97 in RPP21 are not located in the binding interface and are already structured in the free protein, therefore they are not considered to be the candidates for the observed deprotonation. The NMR spectra recorded on the free and RPP21-bound RPP29 indicate that residues 17-48 are only structured in the presence of its partner, including H34 and H46. Though the $pK_a$ of the His side chain is generally determined to be 6, it can certainly shift based on its local chemical environment. For example, the unusual high $pK_a$ of His ($pK_a = 9.7$) has been found in the hemophore HasA.[124] We propose that H34 and H46 may also have higher $pK_a$, resulting in their protonated state in the free protein. Upon binding to RPP21, H34 and H46 are deprotonated due to a change in the local pH, and may allow to form hydrogen bonds using these deprotonated imidazole side chains. Based on the crystal structure of the Pho RPP21-RPP29 complex, H34 and H46 side chains are involved in three hydrogen bonds (Fig. 5.8). Two of them are within RPP29, involving acceptor E47O and protonated donor H34Nε2 (2.74 Å), and deprotonated acceptor H46Nδ1 and acceptor E47N (2.93 Å). One intermolecular hydrogen bond is formed by acceptor RPP21E21Oε2 and protonated donor H46Nε2 (2.86 Å).[97] Due to the limited NOEs, the side chains of H34$^{RPP29}$, E47$^{RPP29}$ and E16$^{RPP21}$ can not be constrained to a stable conformation in our solution structure. Moreover, since we did not know the protonation states of either His when we performed the structure calculation, all His residues in the RPP21-RPP29 were assumed to be protonated on both Nδ1 and Nε2. Thus, Xplor-NIH suite may favor hydrogen-bonding during the structure calculations, so that we might not
observe the hydrogen bonds which requires deprotonated His side chain, or we might see the false hydrogen bonds because of the protonated His side chain. But based on the sequence similarity between Pfu and Pho RPPs, it is possible very similar hydrogen bonding formed in the solution structures of the Pfu RP21-RPP29 complex. However, a caveat of this analysis is that the observed linkage number of two could also arise from partially deprotonating many different functional groups, instead of just those two His residues.

**Thermodynamics of binding coupled folding in the Pfu RPP21-RPP29 complex**

A $\Delta C_p$ predominantly attributed to the protein-protein interaction of the Pfu RPP21-RPP29 complex was obtained after correction for contributions from linked protonation, and yielded a $\Delta C_p$ of 1092.32 ± 161.4 kcal mol$^{-1}$ K$^{-1}$ (Fig. 5.3). This value is still almost twice as large as the predicted one using STC on the assumption of a rigid body interaction between RPP21 and RPP29.

Previously, Spolar and Record showed that the measured $\Delta C_p$ could be used to quantify the extent of coupled binding protein-protein interactions. To do that, they used empirically determined parameters derived from studies of model compounds and protein folding.[85] For an interaction, the $\Delta S_{ass} \text{soc}$ can be decomposed into three elements, entropy changes 1) from the hydrophobic effect ($\Delta S_{\text{HE}}$), 2) due to loss of rotational and transitional freedom ($\Delta S_{rt}$), and 3) from other events including those from binding-induced folding ($\Delta S_{\text{other}}$).

$$\Delta S_{ass \text{soc}} = 0 = \Delta S_{\text{HE}} + \Delta S_{rt} + \Delta S_{\text{other}} \text{(7)}$$
By definition, at the $T_s$, $\Delta S_{assoc}$ is zero, and $\Delta S_{other}$ is balanced by the sum of $\Delta S_{HE}$ and $\Delta S_n$. From transfer of hydrocarbons and amides from water to the pure liquid phase and protein unfolding, it has been shown that $\Delta S_{HE}$ (the entropy change from the hydrophobic effect) is well correlated with change in nonpolar surface area, which can be determined empirically from $\Delta C_p$ in the absence of contribution from other linked events.

$$\Delta S_{HE} = 1.35 \Delta C_p \ln(T_s/386) \quad (8)$$

Studying the rigid body protein-protein interactions, they also found a well-conserved value of $\Delta S_n$ of roughly -50 cal K$^{-1}$ mol$^{-1}$ for a 1:1 binding. Together with the $\Delta C_p$ determined $\Delta S_{HE}$, a structural analysis of protein complexes known to be associated with binding-coupled folding yielded an averaged entropy cost per residue, $\Delta S_{res} = -5.6$ cal K$^{-1}$ mol$^{-1}$. As such, the number of residues folded ($R$, shown as $R$ in Equation 9) can be expressed as:

$$R = \frac{\Delta S_{other}}{\Delta S_{res}} = \frac{\Delta S_{HE} - \Delta S_n}{5.6} = \frac{1.35 \Delta C_p \ln(T_s/386) - 50}{5.6} \quad (9)$$

Following this method, we first determined a $T_s$ of 36.11°C (299.26 K) by fitting the data from the plot of $\Delta G$ vs. temperature to Equation 10 which is rewritten based on the extended Gibbs-Holmholtz equation incorporating $\Delta C_p$ (Fig. 5.9):

$$\Delta G = \Delta H - T\Delta S = \Delta C_p(T - T_n) + T \Delta C_p \ln(T/T_s) \quad (10)$$

where $\Delta C_p$ and $T_n$ come from $\Delta H$ vs. temperature. Then the corrected $\Delta C_p$ (-1092.32 cal mol$^{-1}$ K$^{-1}$) of the Pfu RPP21-RPP29 interaction reveals an estimation of 50 residues that become folded upon binding. This result coincides remarkably well with the folding of 50 residues upon RPP21-RPP29 binding observed in the NMR studies (Chapter 2). These data have allowed us to correlate global thermodynamic results to details of the structural differences between free and bound states of the proteins, and to improve our understanding of the energetics of the RPP21-RPP29 interaction.
CONCLUSION

We characterized the energetics of the *Pfu* RPP21-RPP29 interaction using ITC experiments. A series of titrations performed over a wide range of temperatures allowed us to obtain a $\Delta C_p$ that is twice as expected from rigid body association based on the solution structure of the complex. The ion linkage effect were investigated by carrying out the experiments at different salt concentrations. A slope of 1.4 was observed in the range of 10-150 mM KCl, but was not found to contribute significantly to the measured $\Delta C_p$. In addition, the binding is enhanced at higher salt concentration, perhaps suggesting an important role of ions in preventing the charge repulsion and subsequently facilitating protein folding. Likewise, linked protonation was explored by performing titrations in buffers with different ionization enthalpy at pH 6.7. Two protons are found to be released into the solvent upon binding. Considering the experimental pH and the RPP21-RPP29 structure, this result is consistent with deprotonation of H34 and H46 in RPP29 in order to form hydrogen bonds in the binding interface. After correction for the linked proton equilibrium, a linear fit of $\Delta H$ vs. temperature renders a $\Delta C_p$ of -1092.32 kcal mol$^{-1}$ K$^{-1}$, corresponding to a folding of ~50 residues during complex formation. This is consistent with the observations from the NMR studies. The detailed thermodynamic characterizations improve our understanding the energetics of interaction between RPP21 and RPP29.
Figure 5.2. Enthalpy-entropy compensation in the RPP21-RPP29 interaction. The sign of $\Delta H$ (blue), $-T\Delta S$ (red) change at ~25°C and ~35°C, respectively, resulting in a relatively insensitive $\Delta G$ (black) within the investigated temperature range. The uncertainties are the standard deviations of three measurements repeated under the same condition. Only one experiment was carried out at 20°C.
Figure 5.3. Temperature dependence of the thermodynamics of the RPP21-RPP29 interaction. The $\Delta H$ before and after corrections from proton linkage effects are shown in black and red, respectively. Linear fits of uncorrected and corrected $\Delta H$s report a $\Delta C_p$ of 

$$-1.13348 \pm 0.01616 \text{ kcal mol}^{-1} \text{ K}^{-1}$$ 

and a $T_h$ of 298.51 $\pm$ 4.31 K for the overall reaction, 

and a $\Delta C_p$ of 

$$-1.09232 \pm 0.1614 \text{ kcal mol}^{-1} \text{ K}^{-1}$$ 

and a $T_h$ of 299.84 $\pm$ 4.46 K for the protein-protein interaction within the investigated temperatures. Only a small fraction of $\Delta H$ arises from proton linkage due to the small $\Delta H_i$ of cacodylate buffer, resulting in a small difference in $\Delta C_p$. The reported uncertainties are standard errors from the nonlinear least squares fit of the data to Equation 4.
Figure 5.4. Electrostatic potential map of the Pfu RPP21-RPP29 complex. Electropositivity and electronegativity on the surface are shown in red and blue, respectively. Electrostatic potential maps on the surface of RPP29 (a, left) and RPP21 (a, right) illustrate that the interface is dominated by hydrophobic interactions surrounding prominent charge-charge interactions; the binding partner is shown as a ribbon (RPP21 in cyan and RPP29 in green). (b) Proximity between E47\textsuperscript{RPP29} and R17\textsuperscript{RPP21} and between D72\textsuperscript{RPP29} and R38\textsuperscript{RPP21} in the ensemble of Pfu RPP21-RPP29 complex, (c) Electrostatic potential map of the entire complex reveals two extended electropositive surface patches. The larger of these surfaces (site 1) is composed of residues contributed by both proteins; the smaller (site 2) is localized solely to the RPP21 zinc ribbon.
Figure 5.5. Ion linkage in the *Pfu* RPP21-RPP29 complex. At both 10°C and 55°C, there is ~1.4 ions up taken during the complex formation, indicated by the slope of a linear fit of log $K_A$ vs. log[KCl]. Because data are not linear above 150 mM KCl (blue), the fitting at 10°C only used data in the conditions of 10 mM, 50 mM and 100 mM KCl.
Figure 5.6. Ion linkage and $\Delta C_p$. Temperature dependence of $\Delta H$ was investigated in 10 mM KCl (black, same data shown in red in Fig. 5.3) and 150 mM KCl (red), respectively. Based on the similar slopes ($\Delta C_p = -1127.11 \pm 23.11 \text{ kcal mol}^{-1} \text{ K}^{-1}$ in 150 mM KCl and $\Delta C_p = -1092.32 \pm 16.14 \text{ kcal mol}^{-1} \text{ K}^{-1}$ in 10 mM KCl), ion binding does not contribute significantly to the measured $\Delta C_p$. $\Delta H$ values have been corrected for linked protonation. Uncertainties are from the deviations of the nonlinear squares fit of the integrated heat data to Equation 4.
Figure 5.7. Proton linkage in the *Pfu* RPP21-RPP29 complex. The ΔH values measured in buffers of cacodylate, MOPS and ACES are plotted as a function of the buffer ionization enthalpies. Because the data were acquired at 55°C, the ΔHi of cacodylate (ΔHi = -1.334 kcal mol⁻¹), MOPS (ΔHi = 5.222 kcal mol⁻¹) and ACES (ΔHi = 6.921 kcal mol⁻¹) were adjusted to the values at 55°C (see methods). A linear fit of the data reports a release of two protons upon the RPP21-RPP29 interaction. The errors shown are from the deviations from the nonlinear least squares fit using Equation 4.9. The standard deviations from the linear fitting are shown in the second line for the slope and the intercept, respectively.
Figure 5.8. Hydrogen bonding in the RPP21-RPP29 complex. (a) The crystal structure of the Pho RPP21-RPP29 complex, showing the RPP21 (green)- RPP29 (blue) interface. Residues that fold upon binding are highlighted in bright color (bright green for RPP21, and bright blue in RPP29). The squared area in (a) is zoomed in (b), showing three possible hydrogen bonds (the distance between heteronuclei is less than 3Å). Two of them are intramolecular hydrogen bonds in RPP29 involving E47O (acceptor) and protonated H34Nε2 (donor), and E47N (donor) and H46Nδ1 (acceptor). One intermolecular hydrogen bond is formed between RPP21 E16Oε2 (acceptor) and H46Nε2 (donor). The residue numbers are shown as in Pfu RPP21 and RPP29 based on sequence alignments.
Figure 5.9. $T_s$ obtained from a fit of temperature dependent $\Delta G$ to the $\Delta C_p$ incorporated Gibbs equation. The errors arise from the standard deviations of the nonlinear least square fit using Equation 4. The tightest binding occurs at about 30°C, resulting in a much higher c-value. Therefore, less points were obtained in the transition states, which is used to define $K_A$. Since $\Delta G$ is deduced from $K_A$, the $\Delta G$s reported at 30°C contain the largest errors from the fit.
## APPENDIX A

### LISTS OF INTERMOLECULAR NOES

Table A.1. Intermolecular NOEs identified from chemical shifts in the $^{13}$C-filtered/edited NOESY spectrum recorded on [U-$^{13}$C,$^{15}$N]-RPP29(*) and unlabeled RPP21.\(^5\)

<table>
<thead>
<tr>
<th>RPP29*</th>
<th>RPP21</th>
<th>RPP29*</th>
<th>RPP21</th>
<th>RPP29*</th>
<th>RPP21</th>
</tr>
</thead>
<tbody>
<tr>
<td>S19Hβ</td>
<td>P33Hβ</td>
<td>I23Mδ</td>
<td>V28Hα</td>
<td>I29My2</td>
<td>R27Hδ-</td>
</tr>
<tr>
<td>S19Hβ</td>
<td>L35Mδ-</td>
<td>R26Hδ+</td>
<td>Y31Hε</td>
<td>I29Mδ</td>
<td>V28Hβ</td>
</tr>
<tr>
<td>S19Hβ</td>
<td>L35Mδ+</td>
<td>R26Hδ-</td>
<td>Y31Hε</td>
<td>I29My2</td>
<td>R27Hβ</td>
</tr>
<tr>
<td>S19Hα</td>
<td>E34Hβ1/2</td>
<td>R26Hδ+</td>
<td>Y31Hδ</td>
<td>I29Mδ</td>
<td>L24Hα</td>
</tr>
<tr>
<td>S19Hβ</td>
<td>E34Hβ1/2</td>
<td>R26Hδ+</td>
<td>Y31Hδ</td>
<td>I29My2</td>
<td>V28Hα</td>
</tr>
<tr>
<td>Y20Hα</td>
<td>L35Mδ-</td>
<td>R26Hδ-</td>
<td>Y31Hδ+</td>
<td>I29My2</td>
<td>V28Hβ</td>
</tr>
<tr>
<td>Y20Hα</td>
<td>L35Mδ+</td>
<td>R26Hδ-</td>
<td>Y31Hδ</td>
<td>I29Hα</td>
<td>R27Hδ+</td>
</tr>
<tr>
<td>I23My2</td>
<td>Y31Hβ-</td>
<td>I29Hβ</td>
<td>R27Hγ</td>
<td>I29Mδ</td>
<td>R27Hγ</td>
</tr>
<tr>
<td>I23My2</td>
<td>V28My-</td>
<td>I29Mδ</td>
<td>R27Hα</td>
<td>I29My2</td>
<td>R27Hγ</td>
</tr>
<tr>
<td>I23Mδ</td>
<td>V28My+</td>
<td>I29My2</td>
<td>L24Mδ-</td>
<td>I29My1</td>
<td>V28My+</td>
</tr>
<tr>
<td>I23My2</td>
<td>L35Hβ-</td>
<td>I29My1</td>
<td>R27Hβ</td>
<td>I29Mδ</td>
<td>V28Hα</td>
</tr>
<tr>
<td>I23My2</td>
<td>Y31Hδ</td>
<td>I29My1</td>
<td>Y31Hδ</td>
<td>I29Mδ</td>
<td>L24Mδ-</td>
</tr>
<tr>
<td>I23My1</td>
<td>L35Mδ-</td>
<td>I29Mδ</td>
<td>R27Hβ</td>
<td>F30Hα</td>
<td>L24Mδ-</td>
</tr>
<tr>
<td>I23Mδ</td>
<td>S32Hβ+</td>
<td>I29My2</td>
<td>V28My+</td>
<td>A33Hα</td>
<td>L24Mδ-</td>
</tr>
<tr>
<td>I23Mδ</td>
<td>Y31Hδ</td>
<td>I29My2</td>
<td>R27Hδ+</td>
<td>A33Mβ</td>
<td>L24Mδ+</td>
</tr>
<tr>
<td>I23Mδ</td>
<td>S32Hβ-</td>
<td>I29My2</td>
<td>Y31Hε</td>
<td>A33Mβ</td>
<td>L24Mδ-</td>
</tr>
<tr>
<td>I23My2</td>
<td>Y31Hβ+</td>
<td>I29Mδ</td>
<td>R27Hδ-</td>
<td>A33Hα</td>
<td>R27Hβ</td>
</tr>
<tr>
<td>I23Mδ</td>
<td>Y31Hβ+</td>
<td>I29Hα</td>
<td>R27Hδ-</td>
<td>A33Mβ</td>
<td>R27Hδ+</td>
</tr>
<tr>
<td>I23My2</td>
<td>S32Hβ-</td>
<td>I29Hβ</td>
<td>R27Hβ</td>
<td>A33Mβ</td>
<td>R27Hδ-</td>
</tr>
<tr>
<td>I23Mδ</td>
<td>V28Hβ</td>
<td>I29Mδ</td>
<td>Y31Hε</td>
<td>A33Mβ</td>
<td>L24Hα</td>
</tr>
<tr>
<td>I23My2</td>
<td>L35Hγ</td>
<td>I29My2</td>
<td>L24Hα</td>
<td>A33Mβ</td>
<td>R27Hβ</td>
</tr>
<tr>
<td>RPP29*</td>
<td>RPP21</td>
<td>RPP29*</td>
<td>RPP21</td>
<td>RPP29*</td>
<td>RPP21</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>--------</td>
<td>---------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>I23My2</td>
<td>V28Hβ</td>
<td>I29Mδ</td>
<td>Y31Hδ</td>
<td>H34Hβ+</td>
<td>I20My2</td>
</tr>
<tr>
<td>I23My2</td>
<td>S32Hβ+</td>
<td>I29Mδ</td>
<td>V28My+</td>
<td>H34Hα</td>
<td>I20Hβ</td>
</tr>
<tr>
<td>I23My2</td>
<td>V28Hα</td>
<td>I29My2</td>
<td>Y31Hδ</td>
<td>H34Hα</td>
<td>I20Hα</td>
</tr>
<tr>
<td>H34Hβ-</td>
<td>I20My2</td>
<td>I49Mδ</td>
<td>L42Mδ+</td>
<td>P117Hβ-</td>
<td>L42Mδ-</td>
</tr>
<tr>
<td>H34Hα</td>
<td>I20Mδ</td>
<td>I49Mδ</td>
<td>L24Mδ+</td>
<td>E118Hα</td>
<td>K49Hδ</td>
</tr>
<tr>
<td>H34Hα</td>
<td>I20My2</td>
<td>I49My2</td>
<td>R38Hδ+</td>
<td>E118Hα</td>
<td>L42Mδ+</td>
</tr>
<tr>
<td>N42Hα</td>
<td>I13My2</td>
<td>I49Hβ</td>
<td>L21Mδ-</td>
<td>E118Hα</td>
<td>V46Hα</td>
</tr>
<tr>
<td>N42Hβ+</td>
<td>I13My2</td>
<td>I49My2</td>
<td>Y39Hε</td>
<td>E118Hα</td>
<td>L45Hβ+</td>
</tr>
<tr>
<td>N42Hβ-</td>
<td>I13My2</td>
<td>I49My2</td>
<td>Y39Hδ</td>
<td>E118Hα</td>
<td>L45Mδ-</td>
</tr>
<tr>
<td>W45Hβ+</td>
<td>I13Mδ</td>
<td>I49Mδ</td>
<td>R38Hδ+</td>
<td>L121Mδ-</td>
<td>I13My2</td>
</tr>
<tr>
<td>W45Hα</td>
<td>I13My2</td>
<td>I49Mδ</td>
<td>Y39Hδ</td>
<td>L121Mδ+</td>
<td>I13Mδ</td>
</tr>
<tr>
<td>W45Hβ-</td>
<td>I13Mδ</td>
<td>I49Hβ</td>
<td>L24Mδ+</td>
<td>L121Mδ+</td>
<td>A50Mβ</td>
</tr>
<tr>
<td>W45Hβ-</td>
<td>I13My2</td>
<td>G50Hα+</td>
<td>L24Mδ+</td>
<td>L121Mδ+</td>
<td>L45Mδ-</td>
</tr>
<tr>
<td>W45Hβ+</td>
<td>I13My2</td>
<td>G50Hα+</td>
<td>L24Mδ-</td>
<td>L121Mδ-</td>
<td>L45Mδ-</td>
</tr>
<tr>
<td>H46Hα</td>
<td>I13Mδ</td>
<td>G50Hα-</td>
<td>L24Mδ+</td>
<td>L121Mδ+</td>
<td>K49Hβ+</td>
</tr>
<tr>
<td>E47Hα</td>
<td>R17Hβ</td>
<td>G50Hα-</td>
<td>L24Mδ-</td>
<td>L121Mδ+</td>
<td>K10Hα</td>
</tr>
<tr>
<td>E47Hα</td>
<td>R17Hα</td>
<td>I71My2</td>
<td>Y39Hδ</td>
<td>L121Mδ+</td>
<td>R17Hβ</td>
</tr>
<tr>
<td>E47Hβ</td>
<td>R17Hα</td>
<td>I71Mδ</td>
<td>L35Mδ+</td>
<td>L121Mδ-</td>
<td>V46Hα</td>
</tr>
<tr>
<td>I49My2</td>
<td>L42Mδ+</td>
<td>I71My2</td>
<td>L35Mδ-</td>
<td>L121Mδ-</td>
<td>K10Hα</td>
</tr>
<tr>
<td>I49Mδ</td>
<td>Y39Hε</td>
<td>I71My2</td>
<td>L35Mδ+</td>
<td>L121Mδ-</td>
<td>A50Mβ</td>
</tr>
<tr>
<td>I49Mδ</td>
<td>L21Mδ-</td>
<td>I71Mδ</td>
<td>L35Mδ+</td>
<td>L121Mδ-</td>
<td>R17Hα</td>
</tr>
<tr>
<td>I49My2</td>
<td>L24Hβ</td>
<td>I71Mδ</td>
<td>L35Mδ-</td>
<td>L121Mδ-</td>
<td>A50Hα</td>
</tr>
<tr>
<td>I49My2</td>
<td>L21Mδ-</td>
<td>I71Hβ</td>
<td>L35Mδ-</td>
<td>L121Mδ-</td>
<td>I13Mδ</td>
</tr>
<tr>
<td>I49Mδ</td>
<td>L42Hβ-</td>
<td>I71Hα</td>
<td>L35Mδ+</td>
<td>L121Mδ-</td>
<td>K10Hβ</td>
</tr>
<tr>
<td>I49My2</td>
<td>L42Hβ-</td>
<td>D72Hα</td>
<td>Y39Hδ</td>
<td>L121Mδ+</td>
<td>A50Hα</td>
</tr>
<tr>
<td>I49My2</td>
<td>L24Hα</td>
<td>P117Hβ+</td>
<td>L42Mδ-</td>
<td>L121Mδ+</td>
<td>V46Hα</td>
</tr>
<tr>
<td>I49My2</td>
<td>L24Hγ</td>
<td>P117Hδ+</td>
<td>L42Mδ-</td>
<td>L121Mδ+</td>
<td>R17Hα</td>
</tr>
<tr>
<td>I49My2</td>
<td>L24Mδ+</td>
<td>P117Hγ</td>
<td>L42Mδ-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I49Hα</td>
<td>L24Mδ+</td>
<td>P117Hα</td>
<td>L42Mδ-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5 The downfield and upfield-shifted non-stereoaressed, but spectroscopically distinct diastereotopic methyl groups and methylene protons are indicated by “+” and “-”, respectively.
Table A.2. Intermolecular NOEs identified from chemical shifts in the $^{13}$C-filtered/edited NOESY spectrum recorded on [U-$^{13}$C, $^{15}$N]-RPP21(*) and unlabeled RPP29.

<table>
<thead>
<tr>
<th>RPP21*</th>
<th>RPP29</th>
<th>RPP21*</th>
<th>RPP29</th>
<th>RPP21*</th>
<th>RPP29</th>
</tr>
</thead>
<tbody>
<tr>
<td>K10Hβ</td>
<td>L121Mδ-</td>
<td>I20My2</td>
<td>H34Hβ-</td>
<td>V28My-</td>
<td>I29Hβ</td>
</tr>
<tr>
<td>K10Hα</td>
<td>L121Mδ+</td>
<td>L21Mδ-</td>
<td>I49Hβ</td>
<td>V28Hβ</td>
<td>I29Mδ</td>
</tr>
<tr>
<td>K10Hα</td>
<td>L121Mδ-</td>
<td>L21Mδ-</td>
<td>I49Mγ2</td>
<td>V28My+</td>
<td>I29Hβ</td>
</tr>
<tr>
<td>I13Mγ2</td>
<td>W45Hδ</td>
<td>L21Mδ-</td>
<td>I49Mδ</td>
<td>V28My+</td>
<td>I23Mδ</td>
</tr>
<tr>
<td>I13Mγ2</td>
<td>W45Hβ</td>
<td>L21Mδ+</td>
<td>I49Mδ</td>
<td>V28My+</td>
<td>F30Hζ</td>
</tr>
<tr>
<td>I13Mγ2</td>
<td>H46Hδ2</td>
<td>L21Mδ-</td>
<td>E47Hγ-</td>
<td>V28My-</td>
<td>I29My2</td>
</tr>
<tr>
<td>I13Mδ</td>
<td>W45Hζ3</td>
<td>L21Hα</td>
<td>A33Mβ</td>
<td>V28My-</td>
<td>F30Hε</td>
</tr>
<tr>
<td>I13Mγ2</td>
<td>N42Hα</td>
<td>L21Hβ</td>
<td>A33Mβ</td>
<td>V28My-</td>
<td>I29Mδ</td>
</tr>
<tr>
<td>I13Mγ2</td>
<td>N42Hβ-</td>
<td>L24Mδ+</td>
<td>I49Mδ</td>
<td>V28My+</td>
<td>I23Mγ1</td>
</tr>
<tr>
<td>I13Mγ2</td>
<td>L121Mδ-</td>
<td>L24Mδ-</td>
<td>G50Hα+</td>
<td>V28My+</td>
<td>I29γ1</td>
</tr>
<tr>
<td>I13Mγ2</td>
<td>N42Hβ+</td>
<td>L24Mδ-</td>
<td>F30Hγ+</td>
<td>V28My+</td>
<td>F30Hβ-</td>
</tr>
<tr>
<td>I13Mδ</td>
<td>W45Hε3</td>
<td>L24Mδ-</td>
<td>A33Hα</td>
<td>V28My+</td>
<td>I29Mδ</td>
</tr>
<tr>
<td>I13Mδ</td>
<td>H46Hζ1</td>
<td>L24Mδ+</td>
<td>A33Hα</td>
<td>V28My-</td>
<td>F30Hζ</td>
</tr>
<tr>
<td>A14Hα</td>
<td>L121Mδ-</td>
<td>L24Mδ-</td>
<td>F30Hδ</td>
<td>V28Hα</td>
<td>I29Mδ</td>
</tr>
<tr>
<td>A14Mβ</td>
<td>L121Mδ-</td>
<td>L24Mδ+</td>
<td>F30Hδ</td>
<td>V28Hα</td>
<td>I23Mδ</td>
</tr>
<tr>
<td>A14Mβ</td>
<td>L121Mδ+</td>
<td>L24Mδ-</td>
<td>F30Hα</td>
<td>V28My+</td>
<td>F30Hε</td>
</tr>
<tr>
<td>A14Mβ</td>
<td>L121Hβ-</td>
<td>L24Mδ+</td>
<td>F30Hε</td>
<td>V28My+</td>
<td>I29Mγ2</td>
</tr>
<tr>
<td>A14Hα</td>
<td>L121Mδ-</td>
<td>L24Mδ-</td>
<td>I49My2</td>
<td>V28Hα</td>
<td>I29Mγ2</td>
</tr>
<tr>
<td>A14Mβ</td>
<td>L121Hβ+</td>
<td>L24Mδ-</td>
<td>A33Mβ</td>
<td>V28My-</td>
<td>I23Mγ1</td>
</tr>
<tr>
<td>E16Hγ-</td>
<td>H46Hζ1</td>
<td>L24Mδ-</td>
<td>F30Hε</td>
<td>P30Hγ+</td>
<td>I29Hβ</td>
</tr>
<tr>
<td>E16Hβ-</td>
<td>H46Hζ1</td>
<td>L24Mδ+</td>
<td>G50Hα+</td>
<td>S32Hα</td>
<td>E22Hγ</td>
</tr>
<tr>
<td>E16Hγ+</td>
<td>H46Hδ2</td>
<td>L24Mδ+</td>
<td>A33Mβ</td>
<td>S32Hβ-</td>
<td>I23Mδ</td>
</tr>
<tr>
<td>E16Hγ+</td>
<td>H46Hζ1</td>
<td>L24Mδ+</td>
<td>I49Hβ</td>
<td>S32Hβ-</td>
<td>E22Hβ+</td>
</tr>
<tr>
<td>E16Hβ+</td>
<td>H46Hζ1</td>
<td>L24Mδ+</td>
<td>I49My2</td>
<td>S32Hβ-</td>
<td>E22Hγ</td>
</tr>
<tr>
<td>E16Hα</td>
<td>H46Hζ1</td>
<td>L24Mδ+</td>
<td>G50Hα-</td>
<td>L35Mδ-</td>
<td>I71Mδ</td>
</tr>
<tr>
<td>R17Hα</td>
<td>E47Hβ</td>
<td>L24Mδ-</td>
<td>G50Hα-</td>
<td>L35Mδ-</td>
<td>Y20Hδ</td>
</tr>
<tr>
<td>I20My2</td>
<td>H34Hβ+</td>
<td>L24Mδ+</td>
<td>F30Hβ-</td>
<td>L35Mδ+</td>
<td>I71Mδ</td>
</tr>
<tr>
<td>I20My2</td>
<td>E47Hγ+</td>
<td>R27Hα</td>
<td>I29Hβ</td>
<td>L35Mδ-</td>
<td>I23Hβ</td>
</tr>
<tr>
<td>I20My2</td>
<td>E47Hβ</td>
<td>V28Hβ</td>
<td>I23Mδ</td>
<td>L35Mδ+</td>
<td>I23Hβ</td>
</tr>
<tr>
<td>I20My2</td>
<td>E47Hγ-</td>
<td>V28My-</td>
<td>I29My1</td>
<td>L35Mδ+</td>
<td>I71My2</td>
</tr>
<tr>
<td>I20My2</td>
<td>H34Hδ2</td>
<td>V28My-</td>
<td>I23Mδ</td>
<td>L35Mδ-</td>
<td>Y20Hε</td>
</tr>
</tbody>
</table>
The downfield and upfield-shifted non-stereoassigned, but spectroscopically distinct diastereotopic methyl groups and methylene protons are indicated by “+” and “−”, respectively.

<table>
<thead>
<tr>
<th>RPP21*</th>
<th>RPP29</th>
<th>RPP21*</th>
<th>RPP29</th>
<th>RPP21*</th>
<th>RPP29</th>
</tr>
</thead>
<tbody>
<tr>
<td>L35Mδ-</td>
<td>D72Hβ-</td>
<td>L42Mδ+</td>
<td>I49Mδ</td>
<td>L45Mδ+</td>
<td>P117Hγ</td>
</tr>
<tr>
<td>L35Mδ-</td>
<td>D72Hβ+</td>
<td>L42Mδ+</td>
<td>I49Mγ2</td>
<td>V46Hγ</td>
<td>L121Hα</td>
</tr>
<tr>
<td>L35Mδ+</td>
<td>Y20Hδ</td>
<td>L42Mδ+</td>
<td>P117Hβ-</td>
<td>V46Hγ</td>
<td>L121Mδ-</td>
</tr>
<tr>
<td>L35Mδ-</td>
<td>I23Mγ1</td>
<td>L42Mδ-</td>
<td>P117Hγ</td>
<td>V46Hγ</td>
<td>P117Hβ+</td>
</tr>
<tr>
<td>L35Mδ+</td>
<td>Y20Hε</td>
<td>L42Mδ-</td>
<td>P117Hβ+</td>
<td>V46Hγ</td>
<td>L121Mδ+</td>
</tr>
<tr>
<td>L35Mδ+</td>
<td>F30Hζ</td>
<td>L42Mδ-</td>
<td>I49Mδ</td>
<td>V46Hα</td>
<td>L121Hα</td>
</tr>
<tr>
<td>L35Mδ+</td>
<td>I23Mδ</td>
<td>L42Mδ-</td>
<td>P117Hδ-</td>
<td>V46Hα</td>
<td>L121Mδ+</td>
</tr>
<tr>
<td>L35Mδ+</td>
<td>I23Mγ1</td>
<td>L42Mδ+</td>
<td>P117Hβ+</td>
<td>A50Mβ</td>
<td>L121Mδ+</td>
</tr>
<tr>
<td>L35Mδ+</td>
<td>D72Hβ-</td>
<td>L42Mδ-</td>
<td>E118Hα</td>
<td>A50Mβ</td>
<td>L121Mδ-</td>
</tr>
<tr>
<td>L35Mδ-</td>
<td>I71Hα</td>
<td>L45Mδ+</td>
<td>P117Hδ-</td>
<td>A50Hα</td>
<td>L121Mδ+</td>
</tr>
<tr>
<td>L35Mδ-</td>
<td>I71Mγ2</td>
<td>L45Mδ+</td>
<td>L121Hα</td>
<td>A50Mβ</td>
<td>L121Hβ+</td>
</tr>
<tr>
<td>L42Mδ+</td>
<td>P117Hγ</td>
<td>L45Mδ-</td>
<td>P117Hδ-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L42Mδ-</td>
<td>P117Hβ-</td>
<td>L45Mδ-</td>
<td>P117Hγ</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table A.3. Intermolecular NOEs identified by iterative structure-based assignment (SANE) in $^{13}$C-separated NOESY spectra recorded on [U-$^{13}$C,$^{15}$N]-RPP29(*) and unlabeled RPP21. §

<table>
<thead>
<tr>
<th>RPP29*</th>
<th>RPP21</th>
<th>RPP29*</th>
<th>RPP21</th>
<th>RPP29*</th>
<th>RPP21</th>
</tr>
</thead>
<tbody>
<tr>
<td>S19Hβ</td>
<td>S32Hα</td>
<td>I29My2</td>
<td>R27Hα</td>
<td>W45Hβ-</td>
<td>I13Mδ</td>
</tr>
<tr>
<td>E22Hγ</td>
<td>S32Hβ+</td>
<td>I29My2</td>
<td>V28Hα</td>
<td>W45Hβ+</td>
<td>I13Mδ</td>
</tr>
<tr>
<td>E22Hγ</td>
<td>Y31Hα</td>
<td>I29My2</td>
<td>Y31Hε</td>
<td>W45Hβ+</td>
<td>I13Mδ</td>
</tr>
<tr>
<td>E22Hγ</td>
<td>S32Hα</td>
<td>F30HN</td>
<td>L24Mδ-</td>
<td>H46HN</td>
<td>I13Hβ</td>
</tr>
<tr>
<td>I23Hα</td>
<td>Y31Hβ+</td>
<td>F30Hα</td>
<td>V28My+</td>
<td>H46HN</td>
<td>I13Mγ</td>
</tr>
<tr>
<td>I23My2</td>
<td>V28Hβ</td>
<td>F30Hβ+</td>
<td>L24Mδ-</td>
<td>E47Hα</td>
<td>L24Hβ-</td>
</tr>
<tr>
<td>I23My2</td>
<td>Y31Hβ+</td>
<td>F30Hβ+</td>
<td>L24Mδ+</td>
<td>I49Mδ</td>
<td>L24Hγ</td>
</tr>
<tr>
<td>I23My2</td>
<td>S32Hβ-</td>
<td>G32Hα+</td>
<td>R27Hδ+</td>
<td>I49Mδ</td>
<td>Y39Hδ-</td>
</tr>
<tr>
<td>I23My2</td>
<td>S32Hβ+</td>
<td>A33HN</td>
<td>L24Mδ-</td>
<td>I49Mδ</td>
<td>Y39Hα</td>
</tr>
<tr>
<td>I23My2</td>
<td>S32Hα</td>
<td>A33Hα</td>
<td>L24Mδ-</td>
<td>I49Mδ</td>
<td>Y39Hε</td>
</tr>
<tr>
<td>I23My2</td>
<td>Y31Hδ</td>
<td>A33Mβ</td>
<td>L24Mδ-</td>
<td>I49Mδ</td>
<td>Y39Hδ</td>
</tr>
<tr>
<td>I23Mδ</td>
<td>V28Hβ</td>
<td>A33Mβ</td>
<td>L24Mδ+</td>
<td>I49Mγ2</td>
<td>Y39Hδ-</td>
</tr>
<tr>
<td>I23Mδ</td>
<td>S32Hβ-</td>
<td>A33Mβ</td>
<td>R27Hβ</td>
<td>I49Mγ2</td>
<td>L21Hα</td>
</tr>
<tr>
<td>I23Mδ</td>
<td>S32Hβ+</td>
<td>A33Mβ</td>
<td>R27Hδ-</td>
<td>I49Mγ2</td>
<td>L24Hα</td>
</tr>
<tr>
<td>I23Mδ</td>
<td>Y31Hα</td>
<td>A33Mβ</td>
<td>R27Hδ+</td>
<td>I49Mγ2</td>
<td>Y39Hα</td>
</tr>
<tr>
<td>I23Mδ</td>
<td>S32Hα</td>
<td>A33Mβ</td>
<td>I20Hα</td>
<td>I49Mγ2</td>
<td>Y39Hε</td>
</tr>
<tr>
<td>R26Hβ-</td>
<td>Y31Hβ-</td>
<td>A33Mβ</td>
<td>S23Hα</td>
<td>I49Mγ2</td>
<td>Y39Hδ</td>
</tr>
<tr>
<td>R26Hβ-</td>
<td>Y31Hε</td>
<td>H34HN</td>
<td>I20My2</td>
<td>G50Hα+</td>
<td>L24Mδ-</td>
</tr>
<tr>
<td>R26Hβ-</td>
<td>Y31Hδ</td>
<td>H34Hα</td>
<td>I20Mδ</td>
<td>G50Hα+</td>
<td>L24Mδ+</td>
</tr>
<tr>
<td>I29Hα</td>
<td>R27Hδ+</td>
<td>H34Hβ+</td>
<td>I20My2</td>
<td>G50Hα-</td>
<td>L24Mδ-</td>
</tr>
<tr>
<td>I29Hβ</td>
<td>R27Hγ</td>
<td>H34Hβ-</td>
<td>I20My2</td>
<td>G50Hα-</td>
<td>L24Mδ+</td>
</tr>
<tr>
<td>I29Mδ</td>
<td>Y31Hδ</td>
<td>R35HN</td>
<td>I20Hβ</td>
<td>I71Mγ2</td>
<td>L35Hβ-</td>
</tr>
<tr>
<td>I29Mδ</td>
<td>Y31Hα</td>
<td>R35HN</td>
<td>I20Mδ</td>
<td>I71Mγ2</td>
<td>L35Hγ</td>
</tr>
<tr>
<td>I29Mδ</td>
<td>R27Hα</td>
<td>G36HN</td>
<td>R17Hα</td>
<td>D72Hα</td>
<td>Y39Hδ</td>
</tr>
<tr>
<td>I29Mδ</td>
<td>V28Hα</td>
<td>G36HN</td>
<td>E16Hβ-</td>
<td>D72Hα</td>
<td>L35Mδ-</td>
</tr>
<tr>
<td>I29Mδ</td>
<td>R27Hδ+</td>
<td>G36HN</td>
<td>I20Mδ</td>
<td>E73HN</td>
<td>Y39Hδ</td>
</tr>
<tr>
<td>I29Mδ</td>
<td>Y31Hβ+</td>
<td>G36HN</td>
<td>I20Mγ2</td>
<td>E73HN</td>
<td>L24Hγ</td>
</tr>
<tr>
<td>I29Mδ</td>
<td>V28Hβ</td>
<td>G36Hα+</td>
<td>E16Hβ-</td>
<td>P117Hβ+</td>
<td>L42Mδ-</td>
</tr>
<tr>
<td>I29Mδ</td>
<td>P30Hγ+</td>
<td>G36Hα-</td>
<td>R17Hα</td>
<td>P117Hγ</td>
<td>L42Hβ-</td>
</tr>
<tr>
<td>I29My2</td>
<td>Y31Hβ-</td>
<td>W45HN</td>
<td>I13My2</td>
<td>P117Hγ</td>
<td>L42Mδ-</td>
</tr>
<tr>
<td>I29My2</td>
<td>Y31Hβ+</td>
<td>W45Hβ-</td>
<td>I13My2</td>
<td>E118Hα</td>
<td>V46Hα</td>
</tr>
</tbody>
</table>
The downfield and upfield-shifted non-stereoassigned, but spectroscopically distinct diastereotopic methyl groups and methylene protons are indicated by “+” and “−”, respectively.

<table>
<thead>
<tr>
<th>RPP29*</th>
<th>RPP21</th>
<th>RPP29*</th>
<th>RPP21</th>
<th>RPP29*</th>
<th>RPP21</th>
</tr>
</thead>
<tbody>
<tr>
<td>L121Hβ-</td>
<td>A50Hα</td>
<td>L121Mδ+</td>
<td>V46Hα</td>
<td>L121Mδ-</td>
<td>A14Mβ</td>
</tr>
<tr>
<td>L121Mδ+</td>
<td>L13Mδ+</td>
<td>L121Mδ+</td>
<td>A50Hα</td>
<td>L121Mδ-</td>
<td>A14Hα</td>
</tr>
<tr>
<td>L121Mδ+</td>
<td>I18Hβ</td>
<td>L121Mδ+</td>
<td>A14Hα</td>
<td>K122Hγ-</td>
<td>L45Hβ+</td>
</tr>
<tr>
<td>L121Mδ+</td>
<td>V46Hβ</td>
<td>L121Mδ-</td>
<td>L13Mδ+</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5 The downfield and upfield-shifted non-stereoassigned, but spectroscopically distinct diastereotopic methyl groups and methylene protons are indicated by “+” and “−”, respectively.
Table A.4. Intermolecular NOEs identified by iterative structure-based assignment (SANE) in $^{13}$C-separated NOESY spectra recorded on [U-$^{13}$C,$^{15}$N]-RPP21(*) and unlabeled RPP29. §

<table>
<thead>
<tr>
<th>RPP21*</th>
<th>RPP29</th>
<th>RPP21*</th>
<th>RPP29</th>
<th>RPP21*</th>
<th>RPP29</th>
</tr>
</thead>
<tbody>
<tr>
<td>I13Hα</td>
<td>H46Hα</td>
<td>L21HN</td>
<td>I49Hβ</td>
<td>V28Hα</td>
<td>I29Mγ2</td>
</tr>
<tr>
<td>I13Mγ2</td>
<td>W45Ηδ1</td>
<td>L21Hα</td>
<td>I49Mγ2</td>
<td>V28Mγ-</td>
<td>I29Mδ</td>
</tr>
<tr>
<td>I13Mγ2</td>
<td>N42Hα</td>
<td>L21Mδ+</td>
<td>E47Hγ+</td>
<td>V28Mγ-</td>
<td>F30Ηζ</td>
</tr>
<tr>
<td>I13Mγ2</td>
<td>H46Hα</td>
<td>L21Mδ+</td>
<td>E47Hβ</td>
<td>F29HN</td>
<td>I29Mδ</td>
</tr>
<tr>
<td>I13Mδ</td>
<td>W45Ηδ3</td>
<td>L21Mδ+</td>
<td>E47Hγ-</td>
<td>P30Ηγ+</td>
<td>I29Mδ</td>
</tr>
<tr>
<td>I13Mδ</td>
<td>W45Ηδ1</td>
<td>S23Hα</td>
<td>A33Mβ</td>
<td>Y31HN</td>
<td>I29Mδ</td>
</tr>
<tr>
<td>I13Mδ</td>
<td>W45Ηζ3</td>
<td>S23Hβ</td>
<td>A33Mβ</td>
<td>S32HN</td>
<td>I29Mδ</td>
</tr>
<tr>
<td>I13Mδ</td>
<td>W45Ηβ-</td>
<td>L24Hα</td>
<td>A33Mβ</td>
<td>S32HN</td>
<td>I29Mδ</td>
</tr>
<tr>
<td>I13Mδ</td>
<td>L121Mδ-</td>
<td>L24Mδ-</td>
<td>A33Mβ</td>
<td>S32Hα</td>
<td>E22Hγ</td>
</tr>
<tr>
<td>A14Hα</td>
<td>L121Hβ+</td>
<td>L24Mδ-</td>
<td>F30Hβ+</td>
<td>S32Hβ-</td>
<td>E22Hβ+</td>
</tr>
<tr>
<td>A14Mβ</td>
<td>L121Hβ+</td>
<td>L24Mδ-</td>
<td>H34Hβ-</td>
<td>L35HN</td>
<td>S19Hβ</td>
</tr>
<tr>
<td>E16HN</td>
<td>H46Ηε1</td>
<td>L24Mδ-</td>
<td>I49Hα</td>
<td>L35Mδ+</td>
<td>F30Ηζ</td>
</tr>
<tr>
<td>E16Ηγ+</td>
<td>N42Hβ+</td>
<td>L24Mδ-</td>
<td>G50Hα-</td>
<td>L35Mδ+</td>
<td>I71Hα</td>
</tr>
<tr>
<td>E16Ηγ+</td>
<td>H46Ηε1</td>
<td>L24Mδ-</td>
<td>A33Hα</td>
<td>L35Mδ+</td>
<td>Y20Hα</td>
</tr>
<tr>
<td>E16Ηβ+</td>
<td>H46Ηε1</td>
<td>L24Mδ-</td>
<td>H34Hα</td>
<td>L35Mδ+</td>
<td>I23Mδ</td>
</tr>
<tr>
<td>R17Hα</td>
<td>E47Ηγ-</td>
<td>L24Mδ-</td>
<td>G50Hα+</td>
<td>Y39HN</td>
<td>I49Mδ</td>
</tr>
<tr>
<td>I18HN</td>
<td>E47Ηγ+</td>
<td>L24Mδ-</td>
<td>F30Hδ</td>
<td>L42Mδ-</td>
<td>E73Hβ</td>
</tr>
<tr>
<td>I20HN</td>
<td>E47Ηγ+</td>
<td>L24Mδ-</td>
<td>F30Ηε</td>
<td>L42Mδ-</td>
<td>P117Ηγ</td>
</tr>
<tr>
<td>I20HN</td>
<td>E47Ηβ</td>
<td>L24Mδ-</td>
<td>F30Ηζ</td>
<td>L42Mδ-</td>
<td>P117Ηβ+</td>
</tr>
<tr>
<td>I20Hα</td>
<td>H34Hα</td>
<td>L24Mδ+</td>
<td>F30Hβ-</td>
<td>L42Mδ-</td>
<td>P117Hδ-</td>
</tr>
<tr>
<td>I20Mγ2</td>
<td>H34Ηδ2</td>
<td>L24Mδ+</td>
<td>G50Hα-</td>
<td>L42Mδ-</td>
<td>P117Hβ+</td>
</tr>
<tr>
<td>I20Mγ2</td>
<td>G50Hα+</td>
<td>L24Mδ+</td>
<td>I29Hα</td>
<td>L42Mδ-</td>
<td>R116Ηα</td>
</tr>
<tr>
<td>I20Mγ2</td>
<td>H34Hα</td>
<td>L24Mδ+</td>
<td>G50Hα+</td>
<td>V46HN</td>
<td>L121Hβ+</td>
</tr>
<tr>
<td>I20Mγ2</td>
<td>H34Ηβ-</td>
<td>L24Mδ+</td>
<td>F30Ηε</td>
<td>V46Hα</td>
<td>L121Hβ+</td>
</tr>
<tr>
<td>I20Mγ2</td>
<td>E47Ηβ</td>
<td>R27HN</td>
<td>I23Mδ</td>
<td>V46Hα</td>
<td>L121Mδ+</td>
</tr>
<tr>
<td>I20Mγ2</td>
<td>E47Ηγ-</td>
<td>V28HN</td>
<td>I29Hβ</td>
<td>A50HN</td>
<td>L121Hβ+</td>
</tr>
<tr>
<td>I20Mγ2</td>
<td>I49Ηβ</td>
<td>V28HN</td>
<td>I29Mδ</td>
<td>A50HN</td>
<td>L121Mδ+</td>
</tr>
<tr>
<td>I20Mγ2</td>
<td>A33Mβ</td>
<td>V28HN</td>
<td>I29Mγ2</td>
<td>A50Hα</td>
<td>L121Mδ+</td>
</tr>
<tr>
<td>L21HN</td>
<td>E47Ηγ+</td>
<td>V28Hα</td>
<td>I29Mδ</td>
<td>A50Mβ</td>
<td>L121Mδ+</td>
</tr>
</tbody>
</table>

§ The downfield and upfield-shifted non-stereoassigned, but spectroscopically distinct diastereotopic methyl groups and methylene protons are indicated by "+" and "-".
Figure. A.1. Strips from $^{13}$C-filtered / edited NOE spectra recorded on [U-$^{13}$C, $^{15}$N]RPP29 bound to unlabeled RPP21. The black (positive) cross peaks arise from intermolecular NOEs, while the red (negative) peaks result from incomplete suppression of self- and intramolecular NOEs. The $^1$H and $^{13}$C shifts of the labeled RPP29 are indicated above and below each strip, while the y-axis corresponds to the shift of $^1$Hs from the unlabeled RPP21 to RPP29 proton NOEs.
Figure A.1 continued
Figure. A.2. Strips from $^{13}$C-filtered / edited NOE spectra recorded on [U-$^{13}$C,$^{15}$N]RPP21 bound to unlabeled RPP29. The black (positive) cross peaks arise from intermolecular NOEs, while the red (negative) peaks result from incomplete suppression of self- and intramolecular NOEs. The $^1$H and $^{13}$C shifts of the labeled RPP21 are indicated above and below each strip, while the y-axis corresponds to the shift of $^1$Hs from the unlabeled RPP29 to RPP21 proton NOEs.
Table A.5. Intermolecular NOEs identified from chemical shifts in the \(^{13}\text{C}\)-filtered/edited NOESY spectrum recorded on [U-\(^{13}\text{C},^{15}\text{N}\)]-RPP29Δ17(*) and unlabeled RPP21V14.\(^5\)

<table>
<thead>
<tr>
<th>RPP29Δ17(^*)</th>
<th>RPP21V14</th>
<th>RPP29Δ17(^*)</th>
<th>RPP21V14</th>
<th>RPP29Δ17(^*)</th>
<th>RPP21V14</th>
</tr>
</thead>
<tbody>
<tr>
<td>I29\text{My2}</td>
<td>V28\text{My}+</td>
<td>A33\text{Mβ}</td>
<td>R27\text{Hδ}-</td>
<td>I49\text{My2}</td>
<td>L42\text{Hγ}</td>
</tr>
<tr>
<td>I29\text{My2}</td>
<td>P30\text{Hy}</td>
<td>A33\text{Mβ}</td>
<td>R27\text{Hδ}+</td>
<td>I49\text{My2}</td>
<td>Y39\text{Hδ}</td>
</tr>
<tr>
<td>I29\text{My2}</td>
<td>V28\text{Ha}</td>
<td>A33\text{Mβ}</td>
<td>L24\text{Ha}</td>
<td>I49\text{My2}</td>
<td>Y39\text{Hδ}</td>
</tr>
<tr>
<td>I29\text{My2}</td>
<td>Y31\text{He}</td>
<td>H34\text{Ha}</td>
<td>I20\text{My2}</td>
<td>P117\text{Hβ}-</td>
<td>V46\text{My}+</td>
</tr>
<tr>
<td>I29\text{Mδ}</td>
<td>V28\text{My}+</td>
<td>I49\text{Mδ}</td>
<td>L24\text{Mδ}+</td>
<td>P117\text{Hβ}+</td>
<td>V46\text{My}+</td>
</tr>
<tr>
<td>I29\text{Mδ}</td>
<td>V28\text{Hβ}</td>
<td>I49\text{Mδ}</td>
<td>L42\text{Mδ}+</td>
<td>P117\text{Hβ}+</td>
<td>V46\text{My}+</td>
</tr>
<tr>
<td>I29\text{Mδ}</td>
<td>V28\text{Ha}</td>
<td>I49\text{Mδ}</td>
<td>L42\text{Hγ}</td>
<td>P117\text{Hβ}+</td>
<td>V46\text{My}+</td>
</tr>
<tr>
<td>I29\text{Mδ}</td>
<td>P30\text{Hy}</td>
<td>I49\text{Mδ}</td>
<td>Y39\text{Hδ}</td>
<td>L121\text{Mδ}+</td>
<td>A50\text{Mβ}</td>
</tr>
<tr>
<td>I29\text{Mδ}</td>
<td>Y31\text{He}</td>
<td>I49\text{Mδ}</td>
<td>Y39\text{Hδ}</td>
<td>L121\text{Mδ}+</td>
<td>K49\text{Hδ}</td>
</tr>
<tr>
<td>F30\text{Ha}</td>
<td>L24\text{Mδ}+</td>
<td>I49\text{My2}</td>
<td>L24\text{Mδ}+</td>
<td>L121\text{Mδ}+</td>
<td>K49\text{Hβ}</td>
</tr>
<tr>
<td>A33\text{Mβ}</td>
<td>L24\text{Mδ}+</td>
<td>I49\text{My2}</td>
<td>L24\text{Mδ}-</td>
<td>L121\text{Mδ}+</td>
<td>V46\text{Ha}</td>
</tr>
<tr>
<td>A33\text{Mβ}</td>
<td>R27\text{Hβ}</td>
<td>I49\text{My2}</td>
<td>L42\text{Mδ}+</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^5\) The downfield and upfield-shifted non-stereoassigned, but spectroscopically distinct diastereotopic methyl groups and methylene protons are indicated by “+” and “-”, respectively.
Table A.6. Intermolecular NOEs identified from chemical shifts in the $^{13}$C-filtered/edited NOESY spectrum recorded on [U-$^{13}$C,$^{15}$N]-RPP21V14(*) and unlabeled RPP29Δ17.\(^5\)

<table>
<thead>
<tr>
<th>RPP21V14*</th>
<th>RPP29Δ17</th>
<th>RPP21V14*</th>
<th>RPP29Δ17</th>
<th>RPP21V14*</th>
<th>RPP29Δ17</th>
</tr>
</thead>
<tbody>
<tr>
<td>I20Mδ</td>
<td>H34Hα</td>
<td>V28Hβ</td>
<td>I29My2</td>
<td>S32Hβ+</td>
<td>I23Mδ</td>
</tr>
<tr>
<td>I20My2</td>
<td>H34Hβ-</td>
<td>V28Hβ</td>
<td>I29Mδ</td>
<td>S32Hβ+</td>
<td>I23My2</td>
</tr>
<tr>
<td>I20My2</td>
<td>H34Hβ+</td>
<td>V28My-</td>
<td>I29My2</td>
<td>S32Hβ+</td>
<td>E22Hγ-</td>
</tr>
<tr>
<td>I20My2</td>
<td>H34Hδ2</td>
<td>V28My-</td>
<td>I29Mδ</td>
<td>L35Mδ-</td>
<td>I71My2</td>
</tr>
<tr>
<td>L21Mδ-</td>
<td>I49My2</td>
<td>V28My-</td>
<td>A33Mβ</td>
<td>L35Mδ-</td>
<td>I71Mδ</td>
</tr>
<tr>
<td>L21Mδ-</td>
<td>I49Mδ</td>
<td>V28My+</td>
<td>I29My2</td>
<td>L35Mδ-</td>
<td>I49My1</td>
</tr>
<tr>
<td>L21Mδ-</td>
<td>I49Hβ</td>
<td>V28My+</td>
<td>I29Mδ</td>
<td>L35Mδ-</td>
<td>I71Hβ</td>
</tr>
<tr>
<td>L24Mδ+</td>
<td>I49My2</td>
<td>V28My+</td>
<td>F30Hε</td>
<td>L35Mδ-</td>
<td>I71Hα</td>
</tr>
<tr>
<td>L24Mδ+</td>
<td>A33Mβ</td>
<td>V28My+</td>
<td>F30Hδ</td>
<td>L35Mδ-</td>
<td>F30Hδ</td>
</tr>
<tr>
<td>L24Mδ+</td>
<td>H34Hβ+</td>
<td>Y31Hβ-</td>
<td>I29Mδ</td>
<td>A36Mβ</td>
<td>Y20Hδ</td>
</tr>
<tr>
<td>L24Mδ+</td>
<td>F30Hα</td>
<td>Y31Hβ-</td>
<td>I23Mδ</td>
<td>A36Mβ</td>
<td>Q21Hα</td>
</tr>
<tr>
<td>L24Mδ+</td>
<td>F30Hε</td>
<td>Y31Hβ-</td>
<td>I23My2</td>
<td>L42Mδ+</td>
<td>P117Hβ-</td>
</tr>
<tr>
<td>L24Mδ+</td>
<td>F30Hδ</td>
<td>Y31Hβ+</td>
<td>I29Mδ</td>
<td>L42Mδ+</td>
<td>P117Hβ+</td>
</tr>
<tr>
<td>L24Mδ+</td>
<td>H34Hδ2</td>
<td>Y31Hβ+</td>
<td>I29Mδ</td>
<td>L42Mδ+</td>
<td>P117Hγ+</td>
</tr>
<tr>
<td>L24Mδ-</td>
<td>I29My2</td>
<td>Y31Hβ+</td>
<td>I23My2</td>
<td>L42Mδ+</td>
<td>P117Hδ-</td>
</tr>
<tr>
<td>L24Mδ-</td>
<td>I29Mδ</td>
<td>S32Hα</td>
<td>I23Mδ</td>
<td>L42Mδ+</td>
<td>P117Hδ+</td>
</tr>
<tr>
<td>L24Mδ-</td>
<td>A33Mβ</td>
<td>S32Hα</td>
<td>E22Hβ-</td>
<td>V46Hα</td>
<td>L121Mδ+</td>
</tr>
<tr>
<td>L24Mδ-</td>
<td>H34Hβ+</td>
<td>S32Hα</td>
<td>E22Hγ-</td>
<td>V46Hα</td>
<td>L121Hβ+</td>
</tr>
<tr>
<td>L24Mδ-</td>
<td>F30Hα</td>
<td>S32Hβ-</td>
<td>I23Mδ</td>
<td>V46Hα</td>
<td>P117Hβ-</td>
</tr>
<tr>
<td>L24Mδ-</td>
<td>F30Hε</td>
<td>S32Hβ-</td>
<td>E22Hβ-</td>
<td>V46My+</td>
<td>L121Hβ+</td>
</tr>
<tr>
<td>L24Mδ-</td>
<td>F30Hδ</td>
<td>S32Hβ-</td>
<td>E22Hβ-</td>
<td>V46My+</td>
<td>P117Hβ-</td>
</tr>
<tr>
<td>V28Hα</td>
<td>I29My2</td>
<td>S32Hβ-</td>
<td>E22Hβ+</td>
<td>A50Hα</td>
<td>L121Mδ+</td>
</tr>
<tr>
<td>V28Hα</td>
<td>I29Mδ</td>
<td>S32Hβ-</td>
<td>E22Hγ-</td>
<td>A50Mβ</td>
<td>L121Mδ+</td>
</tr>
</tbody>
</table>

\(^5\) The downfield and upfield-shifted non-stereoassigned, but spectroscopically distinct diastereotopic methyl groups and methylene protons are indicated by “+” and “-”, respectively.
Table A.7. Intermolecular NOEs identified by iterative structure-based assignment (SANE) in $^{13}$C-separated NOESY spectra recorded on [U-$^{13}$C,$^{15}$N]-RPP29Δ17(*) and unlabeled RPP21V14. §

<table>
<thead>
<tr>
<th>RPP29Δ17*</th>
<th>RPP21V14</th>
<th>RPP29Δ17*</th>
<th>RPP21V14</th>
<th>RPP29Δ17*</th>
<th>RPP21V14</th>
</tr>
</thead>
<tbody>
<tr>
<td>I23My2</td>
<td>Y31Hβ+</td>
<td>A33Mβ</td>
<td>R27Hδ-</td>
<td>I49Mδ</td>
<td>L42Hβ</td>
</tr>
<tr>
<td>I23My2</td>
<td>S32Hβ-</td>
<td>A33Mβ</td>
<td>R27Hδ+</td>
<td>I49Mδ</td>
<td>Y39Hα</td>
</tr>
<tr>
<td>I23Mδ</td>
<td>L35Hγ</td>
<td>A33Hα</td>
<td>I20My2</td>
<td>I49Mδ</td>
<td>Y39Hε</td>
</tr>
<tr>
<td>I23Mδ</td>
<td>Y31Hβ-</td>
<td>H34HN</td>
<td>I20My2</td>
<td>I49Mδ</td>
<td>Y39Hδ</td>
</tr>
<tr>
<td>I23Mδ</td>
<td>S32Hα</td>
<td>H34Hα</td>
<td>I20My2</td>
<td>I71Hβ</td>
<td>L35Mδ-</td>
</tr>
<tr>
<td>I29Mδ</td>
<td>V28Hβ</td>
<td>H34Hα</td>
<td>I20Mδ</td>
<td>D72Hβ-</td>
<td>L35Mδ-</td>
</tr>
<tr>
<td>I29Mδ</td>
<td>Y31Hβ-</td>
<td>R35HN</td>
<td>I20Mδ</td>
<td>T74My2</td>
<td>R38Hδ-</td>
</tr>
<tr>
<td>I29Mδ</td>
<td>R27Hδ+</td>
<td>R35HN</td>
<td>I20My2</td>
<td>P117Hγ</td>
<td>L42Mδ+</td>
</tr>
<tr>
<td>I29Mδ</td>
<td>R27Hα</td>
<td>R35HN</td>
<td>I20My2</td>
<td>M119Hα</td>
<td>K49Hδ-</td>
</tr>
<tr>
<td>I29Mδ</td>
<td>V28Hα</td>
<td>G36HN</td>
<td>I20My1</td>
<td>M119Hα</td>
<td>K49Hε-</td>
</tr>
<tr>
<td>I29Mδ</td>
<td>Y31Hα</td>
<td>G36HN</td>
<td>I20My2</td>
<td>L121HN</td>
<td>K49Hδ-</td>
</tr>
<tr>
<td>I29Mδ</td>
<td>Y31Hδ</td>
<td>E47Hγ</td>
<td>L42Mδ-</td>
<td>L121Mδ+</td>
<td>K49Hβ+</td>
</tr>
<tr>
<td>F30HN</td>
<td>L24Mδ-</td>
<td>I49HN</td>
<td>L42Mδ-</td>
<td>L121Mδ+</td>
<td>V46Hα</td>
</tr>
<tr>
<td>F30Hα</td>
<td>L24Hγ</td>
<td>I49My2</td>
<td>Y39Hε</td>
<td>L121Mδ+</td>
<td>A50Hα</td>
</tr>
<tr>
<td>F30Hε</td>
<td>L24Mδ-</td>
<td>I49My2</td>
<td>Y39Hδ</td>
<td>L121Hα</td>
<td>K49Hδ+</td>
</tr>
<tr>
<td>A33HN</td>
<td>R27Hδ-</td>
<td>I49My2</td>
<td>L42Mδ-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A33Mβ</td>
<td>L24Mδ+</td>
<td>I49Mδ</td>
<td>L42Mδ-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

§ The downfield and upfield-shifted non-stereoassigned, but spectroscopically distinct diastereotopic methyl groups and methylene protons are indicated by “+” and “-”, respectively.
Table A.8. Intermolecular NOEs identified by iterative structure-based assignment (SANE) in $^{13}$C-separated NOESY spectra recorded on [U-$^{13}$C, $^{15}$N]-RPP21V14(*) and unlabeled RPP29Δ17. 

<table>
<thead>
<tr>
<th>RPP21V14*</th>
<th>RPP29Δ17</th>
<th>RPP21V14*</th>
<th>RPP29Δ17</th>
<th>RPP21V14*</th>
<th>RPP29Δ17</th>
</tr>
</thead>
<tbody>
<tr>
<td>I20Hα</td>
<td>H34Hα</td>
<td>V28HΝ</td>
<td>I29Mδ</td>
<td>S32Hβ-</td>
<td>E22Hβ+</td>
</tr>
<tr>
<td>I20Mδ</td>
<td>A33Hα</td>
<td>V28Hβ</td>
<td>I29Mδ</td>
<td>S32Hβ-</td>
<td>E22Hα</td>
</tr>
<tr>
<td>I20Mδ</td>
<td>H34Hα</td>
<td>V28Hβ</td>
<td>I29My2</td>
<td>S32Hα</td>
<td>E22Hβ+</td>
</tr>
<tr>
<td>I20My2</td>
<td>H34Hβ-</td>
<td>V28Hα</td>
<td>I29Mδ</td>
<td>L35Mδ-</td>
<td>I71My2</td>
</tr>
<tr>
<td>I20My2</td>
<td>H34Hα</td>
<td>V28Hα</td>
<td>I29My2</td>
<td>L35Mδ-</td>
<td>I71My2</td>
</tr>
<tr>
<td>I20My2</td>
<td>H34Hδ2</td>
<td>V28My-</td>
<td>I29Mδ</td>
<td>L35Mδ+</td>
<td>D72Hβ-</td>
</tr>
<tr>
<td>L24Mδ+</td>
<td>F30Hβ-</td>
<td>V28My-</td>
<td>T27Hα</td>
<td>R38Hδ+</td>
<td>E73Hγ</td>
</tr>
<tr>
<td>L24Mδ+</td>
<td>F30Hβ+</td>
<td>V28My-</td>
<td>F30Hε</td>
<td>R38Hδ+</td>
<td>I49Mδ</td>
</tr>
<tr>
<td>L24Mδ+</td>
<td>I29Hα</td>
<td>Y31HN</td>
<td>I29Mδ</td>
<td>Y39Hν</td>
<td>I49Mδ</td>
</tr>
<tr>
<td>L24Mδ+</td>
<td>F30Hε</td>
<td>Y31Hβ-</td>
<td>I29Mδ</td>
<td>L42Hβ+</td>
<td>I49Mδ</td>
</tr>
<tr>
<td>L24Mδ+</td>
<td>F30Hδ</td>
<td>Y31Hβ-</td>
<td>I23Mδ</td>
<td>L42Mδ-</td>
<td>I49Hβ</td>
</tr>
<tr>
<td>L24Mδ-</td>
<td>H34Hβ-</td>
<td>Y31Hβ+</td>
<td>E22Hβ+</td>
<td>L42Mδ-</td>
<td>P117Hδ-</td>
</tr>
<tr>
<td>L24Mδ-</td>
<td>F30Hβ+</td>
<td>Y31Hβ+</td>
<td>I23Mδ</td>
<td>L42Mδ+</td>
<td>E118Hβ-</td>
</tr>
<tr>
<td>L24Mδ-</td>
<td>A33Mβ</td>
<td>Y31Hε</td>
<td>I29Mδ</td>
<td>L42Mδ+</td>
<td>P117Hγ</td>
</tr>
<tr>
<td>L24Mδ-</td>
<td>I29My2</td>
<td>Y31Hε</td>
<td>R26Hδ-</td>
<td>L42Mδ+</td>
<td>P117Hβ+</td>
</tr>
<tr>
<td>L24Hα</td>
<td>A33Mβ</td>
<td>S32Hβ-</td>
<td>I23My2</td>
<td>L42Mδ+</td>
<td>P117Hδ+</td>
</tr>
<tr>
<td>R27Hδ-</td>
<td>A33Mβ</td>
<td>S32Hβ-</td>
<td>I23My1</td>
<td>L42Mδ+</td>
<td>E118Hα</td>
</tr>
<tr>
<td>R27Hδ-</td>
<td>I29My2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

§ The downfield and upfield-shifted non-stereoassigned, but spectroscopically distinct diastereotopic methyl groups and methylene protons are indicated by “+” and “-”, respectively.
Figure. A.3. Strips from $^{13}$C-filtered / edited NOE spectra recorded on [U-$^{13}$C,$^{15}$N]RPP29Δ17 bound to unlabeled RPP21V14. The black (positive) cross peaks arise from intermolecular NOEs, while the red (negative) peaks result from incomplete suppression of self- and intramolecular NOEs. The $^1$H and $^{13}$C shifts of the labeled RPP29Δ17 are indicated above and below each strip, while the y-axis corresponds to the shift of $^1$Hs from the unlabeled RPP21V14 to RPP29Δ17 proton NOEs.
Figure. A.4. Strips from $^{13}$C-filtered/edited NOE spectra recorded on [U-$^{13}$C,$^{15}$N]RPP21V14 bound to unlabeled RPP29Δ17. The black (positive) cross peaks arise from intermolecular NOEs, while the red (negative) peaks result from incomplete suppression of self- and intramolecular NOEs. The $^1$H and $^{13}$C shifts of the labeled RPP21V14 are indicated above and below each strip, while the y-axis corresponds to the shift of $^1$Hs from the unlabeled RPP29Δ17 to RPP21V14 proton NOEs.
APPENDIX B

BACKBONE ASSIGNMENTS

Pfu RPP29 wild type

Backbone resonance assignments were accomplished using a combination of HNCO, CBCA(CO)NH and HNCACB recorded on U-[\textsuperscript{15}N,\textsuperscript{13}C]-RPP29WT in complex with unlabeled RPP21WT. The following strips are from the HNCACB spectrum, with the chemical shifts of \textsuperscript{1}H, \textsuperscript{15}N at the bottom and the top of the strips, respectively. The \textsuperscript{13}C frequency is at y-axis. Peaks corresponding to C\alpha and C\beta resonances are shown in black and red, respectively. Blue labels indicate the inter- and intra-residual C\alpha and C\beta resonances.
Pfu RPP21 wild type

Backbone resonance assignments were accomplished using a combination of HNCO, CBCA(CO)NH and HNCACB recorded on U-[^{15}N,^{13}C]-RPP21WT in complex with unlabeled RPP29WT. The following strips are from the HNCACB spectrum, with the chemical shifts of $^1$H, $^{15}$N at the bottom and the top of the strips, respectively. The $^{13}$C frequency is at $y$-axis. Peaks corresponding to Cα and Cβ resonances are shown in black and red, respectively. Blue labels indicate the inter- and intra-residual Cα and Cβ resonances.
The representative script (sa_new.inp) is for structure calculations of the *Pfu* RPP21-RPP29 complex using Xplor-NIH suite. The script was initially modified by Carlos Amero from the standard Xplor-NIH script sa.inp. The required input files are labeled in red. The libraries and database supplied by Xplor-NIH suite are indicated in blue. Self-defined parameters are shown in green.

```
remarks file nmr/sa.inp
remarks Simulated annealing protocol for NMR structure determination.
remarks The starting structure for this protocol can be any structure with
remarks a reasonable geometry, such as randomly assigned torsion angles or
remarks extended strands.
remarks Author: Michael Nilges

! Parallel Carlos
evaluate ($numStructs = 50)
evaluate ($randomSeed = 785)

cpyth "from os import environ as env"
cpyth "xplor.command('eval ($proc_num=%s)' % env['XPLOR_PROCESS'])"
cpyth "xplor.command('eval ($num_procs=%s)' % env['XPLOR_NUM_PROCESSES'])"
eval ($num_procs=min($num_procs,$numStructs))
if ( $proc_num >= $num_procs ) then
  stop
end if
```
evaluate ($firstStruct = ($proc_num * $numStructs) / $num_procs)
evaluate ($lastStruct = (($proc_num+1) * $numStructs) / $num_procs)

{====>}
evaluate ($init_t = 5000 )       {*Initial simulated annealing
temperature.*}
{====>}
evaluate ($high_steps= 10000 )   {*Total number of steps at
high temp.*}
{====>}
evaluate ($cool_steps = 6000 )      {*Total number of steps
during cooling.*}

parameter                                        {*Read the
parameter file.*}
{====>}
    @TOPPAR:protein-jx.par
end

{====>}
structure
    @/decay1/joy/X-ray/generate/p29p21wt_ex.psf
end       {*Read the structure file.*}

{====>}
coordinates @/decay1/joy/X-ray/generate/p29p21wt_ex.pdb
{*Read the coordinates.*}
evaluate ($krama = 0.2)
!rgyr
collapse
scale 1.0
!radius of gyration
!ordered region of protein Force, radius
assign (resid 17:424) 50.0 50
end

noe
{====>}
    nres=9000           {*Estimate greater than the actual
number of NOEs.*}
class others
{====>}
    @p21_noe.tbl
    @p29_noe.tbl
    @/decay1/joy/complex/sane/data/p29wt_cara.tbl

188
@/decay1/joy/complex/sane/data/p21wt_cara.tbl
@/decay1/joy/complex/sane/data/p29wt_hbonds.tbl
@/decay1/joy/complex/sane/data/p21wt_hbonds.tbl
class inter
@/decay1/joy/complex/sane/data/cc_cara.tbl
@inter_noe.tbl
diend
{====>}
restraints dihedral
nass = 1000
@/decay1/joy/complex/sane/data/p29wt_talos.tbl
@/decay1/joy/complex/sane/data/p21wt_talos.tbl
{*Read dihedral angle restraints.*}
diend
flags exclude * include bonds angle impr vdw noe cdih xpcs carb rama end

{*Friction c for MD heatbath, in 1/ps.*}
vector do (fbeta=10) (all)
{Uniform heavy masses to speed molecular dynamics.*}
vector do (mass=100) (all)

noe {Parameters for NOE effective energy term.*}
ceiling=1000
averaging * cent
potential * soft
scale * 50.
scale others 50.
scale inter 25.
sqoffset * 0.0
sqconstant * 1.0
sqexponent * 2
soexponent * 1
asymptote * 0.1
rswitch * 0.5
end

carbon
phistep=180
psistep=180
nres=300
class all

189
force 0.5
potential square
@/decay1/joy/p21/sane/lib/rcoil.tbl !rcoil shifts
@/decay1/joy/p21/sane/lib/expected_edited.tbl !13C shift
database
set echo off message off end
set echo on message on end
end

parameter
   {*Parameters for the repulsive energy term.*}
   nbonds
      repel=1.
   rexp=2 irexp=2 rcon=1.
   nbxmod=3
   wmin=0.01
cutnb=4.5 ctonnb=2.99 ctofnb=3.
tolerance=0.5
end
end

restraints dihedral
   scale=5.
end

rama
nres=10000
!set message off echo off end
@/decay1/joy/p21/sane/lib/2D_quarts_new.tbl
@/decay1/joy/p21/sane/lib/3D_quarts_new.tbl
@/4D_quarts_intra_new.tbl
@/decay1/joy/p21/sane/lib/forces_torsion_prot_quarts_intra.tbl
end
set message on echo on end
@/decay1/joy/p21/sane/lib/setup_quarts_torsions_intra_2D3D.tbl
@/setup_quarts_torsions_intra_4D.tbl

{====>}
evaluate ($end_count=25)         {*Loop through a family of 20
structures.*}

coor copy end

evaluate ($count = $firstStruct)
while ($count < $lastStruct ) loop main
   evaluate ($seed = $randomSeed+$count)
set seed $seed end
evaluate ($count=$count+1)

coor swap end
coor copy end

{*
  Initial minimization.*}
restraints dihedral scale=5. end
noe asymptote * 0.1 end
parameter nbonds repel=1. end end
constraints interaction
  (all) (all) weights * 1 vdw 0.002 end end
minimize powell nstep=50 drop=10. nprint=25 end

{*
  High-temperature dynamics.*}
constraints interaction (all) (all)
  weights * 1 angl 0.4 impr 0.1 vdw 0.002 end end
evaluate ($nstep1=int($high_steps * 2. / 3. ) )
evaluate ($nstep2=int($high_steps * 1. / 3. ) )
dynamics verlet
  nstep=$nstep1 timestep=0.003 iasvel=maxwell
  firstt=$init_t
tcoupling=true tbath=$init_t nprint=50 iprfrq=0
end

{*
  Tilt the asymptote and increase weights on geometry.*}
noe asymptote * 1.0 end
constraints interaction
  (all) (all) weights * 1 vdw 0.002 end end
dynamics verlet
  nstep=$nstep2 timestep=0.003 iasvel=current
tcoupling=true
  tbath=$init_t nprint=50 iprfrq=0
end

{*
  Cool the system.*}
restraints dihedral scale=200. end

191
evaluate ($final_t = 100)       { K }
evaluate ($tempstep = 50)       { K }

evaluate ($ncycle = ($init_t-$final_t)/$tempstep)
evaluate ($nstep = int($cool_steps/$ncycle))

evaluate ($ini_rad  = 0.9)        evaluate ($fin_rad  = 0.75)
evaluate ($ini_con=  0.003)       evaluate ($fin_con=  4.0)

evaluate ($bath  = $init_t)
evaluate ($k_vdw = $ini_con)
evaluate ($k_vdwfact = ($fin_con/$ini_con)^(1/$ncycle))
evaluate ($radius=    $ini_rad)
evaluate ($radfact = ($fin_rad/$ini_rad)^(1/$ncycle))

evaluate ($i_cool = 0)
while ($i_cool < $ncycle) loop cool
  evaluate ($i_cool=$i_cool+1)

  evaluate ($bath  = $bath  - $tempstep)
evaluate ($k_vdw=min($fin_con,$k_vdw*$k_vdwfact))
evaluate ($radius=max($fin_rad,$radius*$radfact))

  parameter nbonds repel=$radius   end end
  constraints interaction (all) (all)
    weights * 1. vdw $k_vdw end end

dynamics  verlet
  nstep=$nstep time=0.003 iasvel=current firstt=$bath
tcoup=true tbath=$bath nprint=$nstep iprfreq=0
end

{====>}                                                  {*Abort
condition.*}
evaluate ($critical=$temp/$bath)
if ($critical >  10. ) then
display  ****&&&& rerun job with smaller timestep (i.e.,
0.003)
  stop
end if
end loop cool

{* ================================================= Final
minimization.*}

  constraints interaction (all) (all) weights * 1. vdw 1. end end

192
parameter
  nbonds
    repel=0.80
    rexp=2 irexp=2 rcon=1.
    nbxmod=3
    wmin=0.01
    cutnb=6.0 ctonnb=2.99 ctofnb=3.
    tolerance=1.5
  end
end

minimize powell nstep=1000 drop=10.0 nprint=25 end

{* ============== Write out the final structure(s).*}
print threshold=0.5 noe
evaluate ($rms_noe=$result)
evaluate ($violations_noe=$violations)
print threshold=5. cdih
evaluate ($rms_cdih=$result)
evaluate ($violations_cdih=$violations)
print thres=0.05 bonds
evaluate ($rms_bonds=$result)
print thres=5. angles
evaluate ($rms_angles=$result)
print thres=5. impropers
evaluate ($rms_impropers=$result)
carbon print threshold = 1.0 end
evaluate ($rms_ashift = $rmsca)
evaluate ($rms_bshift = $rmscb)
evaluate ($viol_shift = $violations)
remarks
=================================================================
remarks overall,noe,bonds,angles,improper,vdw,cdih
remarks energies: $ener,$noe, $bond, $angl, $impr, $vdw, $cdih
remarks
=================================================================
remarks bonds,angles,impropers,noe,cdih
remarks rms-d: $rms_bonds,$rms_angles,$rms_impropers,$rms_noe,$rms_cdih
remarks
=================================================================
remarks noe, cdih, pcsi,rdc, shifts
remarks violations.: $violations_noe, $violations_cdih,
$viol_shift
remarks
=================================================================

{=====>}  {*Name(s) of the family of final structures.*}
  evaluate ($filename="anneal_"+encode($count)+".pdb")

    write coordinates output =$filename end
end loop main

stop
APPENDIX D

NMR PULSE SEQUENCES

The pulse sequences of 4D HNCNH NOESY and 3D NCH NOESY were modified from the standard Bruker pulse sequences by Chunhua Yuan.

4D HNCNH NOESY

# 1 "/u/exp/stan/nmr/lists/pp/hn-cn-noesy.cy"
;hn-cn-noesy.cy (November 2007)
;half-filter to select 15N-attached proton in t1
;noesiigpsism3d
;avance-version (00/03/30)
;NOESY-HSQC
;3D sequence with
; homonuclear correlation via dipolar coupling
; dipolar coupling may be due to noe or chemical exchange.
; H-1/X correlation via double inept transfer
; using sensitivity improvement
; simultaneous evolution of C-13 and N-15 chemical shift in t2
;phase sensitive (t1)
;phase sensitive using Echo/Antiecho-TPPI gradient selection (t2)
;with decoupling during acquisition
;using trim pulses in inept transfer
;(use parameterset NOESIIGPSISM3D)
;
;M. Sattler, M. Maurer, J. Schleucher & C. Griesinger,
; J. Biomol. NMR 5, 97-102 (1995)

# 1 "/u/exp/stan/nmr/lists/pp/Avance.incl" 1
;Avance.incl
;
;avance-version (03/02/17)
define list<gradient> EA=<EA>

define delay DELTA
define delay DELTA1
define delay DELTA2
define delay DELTA3
define delay DELTA4
define delay DELTA5
define delay DELTA6
define delay DELTA7
define delay DELTA8

define delay TAU
define delay TAU1
define delay TAU2
define delay TAU3
define delay TAU4
define delay TAU5

;delays for centering pulses

define delay CEN_HN1
define delay CEN_HN2
define delay CEN_HN3
define delay CEN_HC1
define delay CEN_HC2
define delay CEN_HC3
define delay CEN_HC4
define delay CEN_HP1
define delay CEN_HP2
define delay CEN_CN1
define delay CEN_CN2
define delay CEN_CN3
define delay CEN_CN4
define delay CEN_CP1
define delay CEN_CP2

;loop counters

define loopcounter COUNTER
define loopcounter SCALEF
define loopcounter FACTOR1
define loopcounter FACTOR2
define loopcounter FACTOR3

;$Id: Delay.incl,v 1.11 2002/06/12 09:04:22 ber Exp $
#$ 24 "/u/exp/stan/nmr/lists/pp/hn-cn-noesy.cy" 2

"p2=p1*2"
"p4=p3*2"
"p22=p21*2"
"d4=1s/(cnst2*4)"
"d0=3u"
"d10=3u"
"d11=30m"
"d12=20u"
"d13=3u"
"d20=3u"
"d26=1s/(cnst4*4)"

"d21=p16+d16-7u"
"d24=d4/2"

"DELTA1=d26-d4-p4+d13"
"DELTA2=p16+d16+d20*2+p2-p3*2+p4+d10*4+p2*2+p1*6+15u"
"DELTA3=p16+d16+d10*2+p2+p1*2+6u"
"DELTA4=d4*0.5+p21+p1-p2-p22"
"DELTA5=d26-d4*0.5-p2-p22-p3-p4-(p21*0.27)"
"DELTA6=(p21*0.27)-(p1*0.27)"
"DELTA7=d26-d4*0.5"
"DELTA8=d26-p21*2-p8-d0*2-3u"
"CEN_HN2=(p22-p2)/2"
"CEN_CN2=(p8-p6)/2"

aqseq 321

# 1 "mc_line 69 file /u/exp/stan/nmr/lists/pp/hn-cn-noesy.cy
expanding definition part of mc command before ze"
; dimension 3 aq-mode (F2) Echo-Antiecho (F1) States-TPPI  F2->F1
define delay MCWRK
define delay MCREST
define loopcounter ST1CNT
"ST1CNT = td1 / (2)"
define loopcounter ST2CNT
"ST2CNT = td2 / (2)"
"MCWRK = 0.071429*d1"
"MCREST = d1 - d1"
# 69 "/u/exp/stan/nmr/lists/pp/hn-cn-noesy.cy"
1 ze
# 1 "mc_line 69 file /u/exp/stan/nmr/lists/pp/hn-cn-noesy.cy
expanding definition of mc command after ze"
# 70 "/u/exp/stan/nmr/lists/pp/hn-cn-noesy.cy"
   d11 pl12:f2 pl16:f3
# 1 "mc_line 71 file /u/exp/stan/nmr/lists/pp/hn-cn-noesy.cy
expanding start label for mc command"
2 MCWRK * 3 do:f2 do:f3
LBLSTS2, MCWRK * 7
LBLF2, MCWRK * 3
LBLSTS1, MCWRK
LBLF1, MCREST
# 72 "/u/exp/stan/nmr/lists/pp/hn-cn-noesy.cy"
3 d12 pl2:f2 pl3:f3
(p1 ph12)
;--- half filter ---------------
d26
(p2 ph12)
(p21 ph23):f3
3u
(p21 ph24):f3
DELTA8 p15:f2 p16:f3
;-----------------------------
d0
(CEN_CN2 p6 ph1):f2 (p8 ph1):f3 ;(p15,p6):f2 and
(p16,p8):f3 for simultaneous
(d0
(p1 ph1) ;for simultaneous 13C/15N pulses
(d0
(d8*0.5
(CEN_CN2 p6 ph1):f2 (p8 ph1):f3 ;both p15 and p16 +3dB
from hard pulse
(d8*0.5
(p1 ph1)
d4 pl2:f2 pl3:f3
(p4 ph6):f2
DELTA1
(CEN_HN2 p2 ph1) (p22 ph7):f3
d26 setnmr2|0 setnmr0|34|32|33
p28 ph1
d13
(p1 ph2)
3u
p16:gp1
d16
(p21 ph8):f3
3u
p16:gp2*EA
d16
DELTA2
(p22 ph8):f3
d20
(p1 ph2 3u p2 ph1 3u p1 ph2)
d20
(p3 ph3):f2
d10
(p1 ph1 3u p2 ph2 3u p1 ph1)
d10
p16:gp3*EA
d16
(p4 ph4):f2
DELTA3
(p3 ph4):f2
(p21 ph10):f3
(p1 ph1)
d24
(p4 ph1):f2
(p2 ph1)
(p22 ph1):f3
DELTA4
(p3 ph5):f2
DELTA7
(p2 ph1)
(p22 ph1):f3
DELTA5
(p21 ph11):f3
DELTA6
(p1 ph2)
d4
(p4 ph1):f2
DELTA1
(CEN_HN2 p2 ph1) (p22 ph1):f3
d26
d13
(p1 ph1)
d21
(p1 ph2 3u p2 ph1 3u p1 ph2)
3u
p16:gp4
d16 pl12:f2 pl16:f3
4u setnmr2^0 setnmr0^34^32^33
go=2 ph31 cpd2:f2 cpd3:f3
# 1 "mc_line 148 file /u/exp/stan/nmr/lists/pp/hn-cn-noesy.cy
expanding mc command in line"
MCWRK do:f2 do:f3 wr #0 if #0 zd igrad EA MCWRK ip5*2 MCWRK
ip11*2
lo to LBLSTS2 times 2
MCWRK id10 MCWRK id20 MCWRK ip3*2 MCWRK ip6*2 MCWRK
ip7*2 MCWRK ip8*2 MCWRK ip31*2
lo to LBLF2 times ST2CNT
MCWRK rd10 MCWRK rd20 MCWRK ip12
lo to LBLSTS1 times 2
MCWRK id0
lo to LBLF1 times ST1CNT
# 151 "/u/exp/stan/nmr/lists/pp/hn-cn-noesy.cy"
exit

ph1=0
ph2=1
ph3=0 0 2 2
ph4=0 0 0 2 2 2 2
ph5=3 3 3 1 1 1 1

200
ph6=0
ph7=0
ph8=2 2 0 0
ph10=0 0 0 0 2 2 2 2
ph11=1 1 1 1 3 3 3 3
ph12=0 0 0 0 0 0 0 0
   2 2 2 2 2 2 2 2
ph23=0 0
ph24=0 2
ph31=0 2 2 0 2 0 0 2
       2 0 0 2 0 2 2 0

;p11 : f1 channel - power level for pulse (default)
p12 : f2 channel - power level for pulse (default)
p13 : f3 channel - power level for pulse (default)
p112: f2 channel - power level for CPD/BB decoupling
p116: f3 channel - power level for CPD/BB decoupling
p1 : f1 channel - 90 degree high power pulse
p2 : f1 channel - 180 degree high power pulse
p3 : f2 channel - 90 degree high power pulse
p4 : f2 channel - 180 degree high power pulse
p16: homospoil/gradient pulse
p28: f1 channel - trim pulse
p21: f3 channel - 90 degree high power pulse
p22: f3 channel - 180 degree high power pulse

;d0 : incremented delay (F1 in 3D)                  [3 usec]
d1 : relaxation delay; 1-5 * T1
;d4 : 1/(4J(CH))
d8 : mixing time
;d10: incremented delay (F2 in 3D)                 [3 usec]
d11: delay for disk I/O                             [30 msec]
d12: delay for power switching                      [20 usec]
d13: short delay                                     [3 usec]
d16: delay for homospoil/gradient recovery
;d20: incremented delay (F2 in 3D)                 [3 usec]
d21: = p16 + d16 + 7u
;d24: = d4 / 2
;d26: 1/(4J(NH))

;DELTA1: d26-d4-p4+d13
;DELTA2: p16+d16+d20*2+p2+p3*2+p4+d10*4+p2*2+p1*6+15u
;DELTA3: p16+d16+d10*2+p2+p1*2+6u
;DELTA4: d4*0.5+p21+p1-p2-p22
;DELTA5: d26-d4*0.5-p2-p22+p3-p4-(p21*0.27)
;DELTA6: (p21*0.27)-(p1*0.27)
;DELTA7:  d26-d4*0.5
;DELTA8:  d0*2+p22
;CEN_HN2:  (p22-p2)/2
;CEN_CN2:  (p22-p4)/2

;cnst2:  = J(XH)
;cnst4:  = J(YH)
;in0:  1/(2 * SW(H)) = DW(H)
;nd0:  2
;in10:  1/(2 * SW(C)) = DW(C)
;nd10:  2
;in20:  in20 + in10 = 1/(2 * SW(N)) = DW(N)
;NS:  8 * n
;DS:  >= 16
;td1:  number of experiments in F1
;td2:  number of experiments in F2
;FnMODE: States-TPPI (or TPPI) in F1
;FnMODE: echo-antiecho in F2
;cpd2:  decoupling according to sequence defined by cpdprg2
;cpd3:  decoupling according to sequence defined by cpdprg3
;pcpd2:  f2 channel - 90 degree pulse for decoupling sequence
;pcpd3:  f3 channel - 90 degree pulse for decoupling sequence

;use gradient ratio:    gp 1 :  gp 2 :  gp 3 : gp 4
;                    80 : -47.63 : 32.37 : 8.14

;for z-only gradients:
gpz1:  80%
gpz2:  -47.63%
gpz3:  32.37%
gpz4:  8.14%

;use gradient files:
gpnam1:  SINE.100
gpnam2:  SINE.100
gpnam3:  SINE.100
gpnam4:  SINE.100

;for older datasets use AQORDER : 3 - 2 - 1

;$Id: noesiigpsism3d,v 1.7 2000/05/08 11:40:47 eng Exp$
3D NCH NOESY

# 1 "/u/exp/stan/nmr/lists/pp/noesyncgp3d"
;noesyncgp3d
;avance-version (03/08/05)
;NOESY-HSQC
;3D sequence with
; homonuclear correlation via dipolar coupling
; dipolar coupling may be due to noe or chemical exchange.
; H-1/X correlation via double inept transfer
; using sensitivity improvement
;
; F1(H) -> F3(N,t1) -> F1(H) -> F2(C,t2) -> F1(H,t3)
;
;phase sensitive (t1)
;phase sensitive using Echo/Antiecho-TPPI gradient selection (t2)
;with decoupling during acquisition
;using shaped pulses for inversion on f2 - channel
;(use parameterset NOESYNCGP3D)
;
;T. Diercks, M. Coles & H. Kessler, J. Biomol. NMR, 15,
; 177-180 (1999)

prosol relations=<triple>

# 1 "/u/exp/stan/nmr/lists/pp/Avance.incl" 1
;Avance.incl
;
;avance-version (03/02/17)

;$Id: Avance1.incl,v 1.7.2.3 2003/02/25 14:48:47 ber Exp $
# 25 "/u/exp/stan/nmr/lists/pp/noesyncgp3d" 2

# 1 "/u/exp/stan/nmr/lists/pp/Grad.incl" 1
;Grad.incl - include file for Gradient Spectroscopy
;
;avance-version (02/05/31)

define list<gradient> EA=<EA>

;$Id: Grad1.incl,v 1.7 2002/06/12 09:04:22 ber Exp $
# 26 "/u/exp/stan/nmr/lists/pp/noesyncgp3d" 2
;Delay.incl - include file for commonly used delays
;
;version 00/02/07

; general delays
define delay DELTA
define delay DELTA1
define delay DELTA2
define delay DELTA3
define delay DELTA4
define delay DELTA5
define delay DELTA6
define delay DELTA7
define delay DELTA8

define delay TAU
define delay TAU1
define delay TAU2
define delay TAU3
define delay TAU4
define delay TAU5

; delays for centering pulses
define delay CEN_HN1
define delay CEN_HN2
define delay CEN_HN3
define delay CEN_HC1
define delay CEN_HC2
define delay CEN_HC3
define delay CEN_HC4
define delay CEN_HP1
define delay CEN_HP2
define delay CEN_CN1
define delay CEN_CN2
define delay CEN_CN3
define delay CEN_CN4
define delay CEN_CP1
define delay CEN_CP2

; loop counters
define loopcounter COUNTER
define loopcounter SCALEF
define loopcounter FACTOR1
define loopcounter FACTOR2
define loopcounter FACTOR3

;p2=p1*2" 
"p4=p3*2"
"p22=p21*2"
"d0=3u"
"d4=1s/(cnst2*4)"
"d11=30m"
"d13=4u"
"d26=1s/(cnst4*4)"

"DELTA1=larger(p2,p8)+d0*2"
"DELTA2=d4-larger(p2,p8)/2"
"DELTA3=p16+d16+d10*2"
"DELTA4=d13+p16+d16+4u"
"TAU=d8-p16-d16"

"spoff5=bf2*(cnst21/1000000)-o2"
"spoff13=bf2*(cnst26/1000000)-o2"

aqseq 321

# 1 "mc_line 63 file /u/exp/stan/nmr/lists/pp/noesyncgp3d expanding definition part of mc command before ze" 
; dimension 3 aq-mode (F2) Echo-Antiecho (F1) States-TPPI F2->F1 define delay MCWRK define delay MCREST
define loopcounter ST1CNT
"ST1CNT = td1 / (2)"
define loopcounter ST2CNT
"ST2CNT = td2 / (2)"
"MCWRK = 0.111111*d1"
"MCREST = d1 - d1"
# 63 "/u/exp/stan/nmr/lists/pp/noesyncgp3d"
1 ze
# 1 "mc_line 63 file /u/exp/stan/nmr/lists/pp/noesyncgp3d
expanding definition of mc command after ze"
# 64 "/u/exp/stan/nmr/lists/pp/noesyncgp3d"
d11 pl12:f2
# 1 "mc_line 65 file /u/exp/stan/nmr/lists/pp/noesyncgp3d
expanding start label for mc command"
2 MCWRK * 2 do:f2
LBLSTS2, MCWRK * 4
LBLF2, MCWRK * 2
LBLSTS1, MCWRK
LBLF1, MCREST
# 66 "/u/exp/stan/nmr/lists/pp/noesyncgp3d"
3 (p1 ph1)
d26
(center (p2 ph2) (p22 ph7):f3 )
d26 setnmr2|0 setnmr0|34|32|33
(p1 ph2)
p16:gp1
d16
(p21 ph3):f3
do0
(center (p2 ph8) (p8:sp13 ph1):f2 )
do0
(p22 ph1):f3
DELTAl
(p21 ph1):f3
p16:gp1
d16
(p1 ph2)
d26
(center (p2 ph2) (p22 ph1):f3 )
d26
(p1 ph9)
TAU
p16:gp2
d16
(p1 ph1)
DELTA2
(center (p2 ph1) (p8:sp13 ph10):f2 )
DELTA2
(p28 ph1)
4u
(p1 ph2)

p16:gp3
d16 pl2:f2

(p3 ph4):f2
2u
d10
(center (p2 ph8) (p14:sp5 ph1):f2 (p22 ph1):f3 )
d10
2u
p16:gp4*EA
d16
(p4 ph5):f2
DELTA3
(p14:sp5 ph1):f2
4u

(center (p1 ph1) (p3 ph5):f2 )
d24
(center (p2 ph2) (p4 ph1):f2 )
d24
(center (p1 ph2) (p3 ph6):f2 )
DELTA2
(center (p2 ph2) (p8:sp13 ph1):f2 )
DELTA2
(p1 ph1)
DELTA4
(p2 ph1)
d13
p16:gp5
d16 pl12:f2
4u setnmr2^0 setnmr0^34^32^33

# 1 "mc_line 136 file /u/exp/stan/nmr/lists/pp/nosyncgp3d
expanding mc command in line"
MCWRK do:f2 wr #0 if #0 zd igrad EA MCWRK ip6*2
lo to LBLSTS2 times 2
MCWRK id10 MCWRK ip4*2 MCWRK ip10*2 MCWRK ip31*2
lo to LBLF2 times ST2CNT
MCWRK ip3 MCWRK rd10
lo to LBLSTS1 times 2

207
MCWRK id0
lo to LBLF1 times ST1CNT
# 139 "/u/exp/stan/nmr/lists/pp/noesyncgp3d"
exit

ph1=0
ph2=1
ph3=0 0 2 2
ph4=0 2
ph5=0 0 0 2 2 2 2
ph6=3 3 3 1 1 1 1
ph7=0
ph8=1 1 1 1 1 1 1 3 3 3 3 3 3 3
ph9=2
ph10=0
ph31=0 2 2 0 2 0 2

;p11 : f1 channel - power level for pulse (default)
p12 : f2 channel - power level for pulse (default)
p13 : f3 channel - power level for pulse (default)
p12: f2 channel - power level for CPD/BB decoupling
sp5: f2 channel - shaped pulse 180 degree (C=O off resonance)
sp13: f2 channel - shaped pulse 180 degree (adiabatic)
p1 : f1 channel - 90 degree high power pulse
p2 : f1 channel - 180 degree high power pulse
p3 : f2 channel - 90 degree high power pulse
p4 : f2 channel - 180 degree high power pulse
p8 : f2 channel - 180 degree shaped pulse for inversion (adiabatic)
p14: f2 channel - 180 degree shaped pulse
p16: homospoil/gradient pulse [1 msec]
p21: f3 channel - 90 degree high power pulse
p22: f3 channel - 180 degree high power pulse
p28: f1 channel - trim pulse [1 msec]
d0 : incremented delay (F1 in 3D) [3 usec]
d1 : relaxation delay; 1-5 * T1
d4 : 1/(4J)CH
d8 : mixing time
d10: incremented delay (F2 in 3D) [3 usec]
d11: delay for disk I/O [30 msec]
d13: short delay [4 usec]
d16: delay for homospoil/gradient recovery
d24: 1/(4J)CH for CH
; 1/(8J)CH for all multiplicities
d26: 1/(4J(NH))
;cnst2: = J(CH)
;cnst4: = J(NH)
;cnst21: CO chemical shift (offset, in ppm)
cnst26: Call chemical shift (offset, in ppm) [101 ppm]
in0: 1/(2 * SW(N)) = DW(N)
nd0: 2
in10: 1/(2 * SW(C)) = DW(C)
nd10: 2
NS: 8 * n
DS: >= 32
;td1: number of experiments in F1
td2: number of experiments in F2
;FnMODE: States-TPPI (or TPPI) in F1
;FnMODE: echo-antiecho in F2
cpd2: decoupling according to sequence defined by cpdprg2
;pcpd2: f2 channel - 90 degree pulse for decoupling sequence

;use gradient ratio: gp 1 : gp 2 : gp 3 : gp 4 : gp 5
5 : 50 : 30 : 80 : 20.1

;for z-only gradients:
gpz1: 5%
gpz2: 50%
gpz3: 30%
gpz4: 80%
gpz5: 20.1%

;use gradient files:
gpnam1: SINE.100
gpnam2: SINE.100
gpnam3: SINE.100
gpnam4: SINE.100
gpnam5: SINE.100

;$Id: noesyncgp3d,v 1.2.2.2 2003/08/05 13:12:37 ber Exp $
APPENDIX E

ITC EXPERIMENTS

The ITC experiments were performed by injecting Pfu RPP29WT into Pfu RPP21WT, or vice versa. The reported thermodynamic parameters (stoichiometry n, $K_A$, $\Delta H$ and $\Delta S$) are summed in the following tables. The errors are from the nonlinear least square fit of the data to a one binding site model. Representative isotherms are shown for experiments performed in each condition.
Table E.1. Table of the thermodynamic parameters measured to obtain $\Delta C_p$ in 10 mM KCl. *

<table>
<thead>
<tr>
<th>Temp (°C)</th>
<th>Run</th>
<th>Stoichiometry</th>
<th>$K_a$</th>
<th>$\Delta H$</th>
<th>$\Delta S$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$n$</td>
<td>($10^6$ M$^{-1}$)</td>
<td>(kcal mol$^{-1}$)</td>
<td>(cal mol$^{-1}$ K$^{-1}$)</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>0.894 ± 0.004</td>
<td>8.47 ± 0.84</td>
<td>17.25 ± 0.12</td>
<td>92.6</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>1.02 ± 0.005</td>
<td>7.34 ± 0.85</td>
<td>17.54 ± 0.15</td>
<td>93.3</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>0.954 ± 0.003</td>
<td>11.8 ± 1.03</td>
<td>16.51 ± 0.09</td>
<td>90.6</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1.02 ± 0.007</td>
<td>9.82 ± 1.81</td>
<td>11.81 ± 0.14</td>
<td>73.0</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>1.08 ± 0.008</td>
<td>9.71 ± 1.79</td>
<td>11.70 ± 0.14</td>
<td>72.5</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>0.944 ± 0.007</td>
<td>13.2 ± 2.71</td>
<td>11.41 ± 0.14</td>
<td>72.2</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>0.900 ± 0.014</td>
<td>9.55 ± 3.48</td>
<td>5.95 ± 0.15</td>
<td>52.2</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>1.03 ± 0.010</td>
<td>36.3 ± 16.7</td>
<td>-4.87 ± 0.10</td>
<td>18.5</td>
</tr>
<tr>
<td>30</td>
<td>2</td>
<td>0.936 ± 0.009</td>
<td>133 ± 117</td>
<td>-4.56 ± 0.11</td>
<td>22.1</td>
</tr>
<tr>
<td>30</td>
<td>3</td>
<td>0.941 ± 0.014</td>
<td>30.0 ± 18.4</td>
<td>-5.51 ± 0.15</td>
<td>16.1</td>
</tr>
<tr>
<td>35</td>
<td>1</td>
<td>0.898 ± 0.004</td>
<td>22.9 ± 3.50</td>
<td>-10.10 ± 0.08</td>
<td>0.9</td>
</tr>
<tr>
<td>35</td>
<td>2</td>
<td>0.995 ± 0.006</td>
<td>17.8 ± 3.81</td>
<td>-10.11 ± 0.11</td>
<td>0.4</td>
</tr>
<tr>
<td>35</td>
<td>3</td>
<td>1.10 ± 0.005</td>
<td>19.6 ± 3.40</td>
<td>-10.54 ± 0.09</td>
<td>-0.8</td>
</tr>
<tr>
<td>40</td>
<td>1</td>
<td>1.05 ± 0.005</td>
<td>21.9 ± 3.95</td>
<td>-14.52 ± 0.13</td>
<td>-12.8</td>
</tr>
<tr>
<td>40</td>
<td>2</td>
<td>1.03 ± 0.004</td>
<td>15.0 ± 1.95</td>
<td>-17.05 ± 0.13</td>
<td>-21.6</td>
</tr>
<tr>
<td>40</td>
<td>3</td>
<td>1.01 ± 0.004</td>
<td>14.4 ± 1.89</td>
<td>-15.96 ± 0.12</td>
<td>-18.2</td>
</tr>
<tr>
<td>45</td>
<td>1</td>
<td>0.936 ± 0.003</td>
<td>11.6 ± 0.99</td>
<td>-23.92 ± 0.13</td>
<td>-42.9</td>
</tr>
<tr>
<td>45</td>
<td>2</td>
<td>1.01 ± 0.004</td>
<td>10.5 ± 1.15</td>
<td>-23.59 ± 0.17</td>
<td>-42.0</td>
</tr>
<tr>
<td>45</td>
<td>3</td>
<td>1.06 ± 0.004</td>
<td>9.87 ± 1.04</td>
<td>-23.08 ± 0.16</td>
<td>-40.5</td>
</tr>
</tbody>
</table>

* All experiments were carried out in the standard ITC buffer (20 mM cacodylate, pH 6.7, 10 mM KCl, 0.3 mM ZnCl$_2$ and 0.02% NaN$_3$). Titrations of RPP29 (~200 μM) into RPP21 (~20 μM) were performed at different temperatures ranging from 10°C to 45°C.
Table E.2. Table of the thermodynamic parameters measured at different salt concentrations at 55°C. *

<table>
<thead>
<tr>
<th>[KCl] (mM)</th>
<th>Run</th>
<th>Stoichiometry</th>
<th>$K_A$ $^b$ ($10^7$ M$^{-1}$)</th>
<th>$\Delta H$ (kcal mol$^{-1}$)</th>
<th>$\Delta S$ (cal mol$^{-1}$ K$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10$^a$</td>
<td>1</td>
<td>0.959 ± 0.004</td>
<td>0.161 ± 0.007</td>
<td>-40.62 ± 0.24</td>
<td>95.4</td>
</tr>
<tr>
<td>10$^a$</td>
<td>2</td>
<td>0.933 ± 0.005</td>
<td>0.159 ± 0.009</td>
<td>-40.30 ± 0.30</td>
<td>-94.4</td>
</tr>
<tr>
<td>10$^a$</td>
<td>3</td>
<td>0.950 ± 0.003</td>
<td>0.195 ± 0.008</td>
<td>-37.92 ± 0.18</td>
<td>-86.8</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>0.943 ± 0.003</td>
<td>1.73 ± 0.15</td>
<td>-39.83 ± 0.19</td>
<td>-88.3</td>
</tr>
<tr>
<td>50</td>
<td>2</td>
<td>0.952 ± 0.003</td>
<td>1.76 ± 0.15</td>
<td>-38.16 ± 0.18</td>
<td>-83.1</td>
</tr>
<tr>
<td>50</td>
<td>3</td>
<td>0.943 ± 0.003</td>
<td>1.91 ± 0.18</td>
<td>-38.39 ± 0.19</td>
<td>-83.7</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>1.12 ± 0.004</td>
<td>3.87 ± 0.59</td>
<td>-31.63 ± 0.20</td>
<td>-61.7</td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>0.973 ± 0.002</td>
<td>4.73 ± 0.61</td>
<td>-35.82 ± 0.18</td>
<td>-74.0</td>
</tr>
<tr>
<td>100</td>
<td>3</td>
<td>0.996 ± 0.002</td>
<td>4.38 ± 0.40</td>
<td>-34.50 ± 0.12</td>
<td>-70.2</td>
</tr>
<tr>
<td>150</td>
<td>1</td>
<td>1.07 ± 0.005</td>
<td>7.71 ± 2.1</td>
<td>-32.07 ± 0.30</td>
<td>-61.6</td>
</tr>
<tr>
<td>150</td>
<td>2</td>
<td>1.01 ± 0.002</td>
<td>7.72 ± 1.2</td>
<td>-31.85 ± 0.16</td>
<td>-61.0</td>
</tr>
<tr>
<td>150</td>
<td>3</td>
<td>0.869 ± 0.003</td>
<td>7.26 ± 1.6</td>
<td>-34.80 ± 0.27</td>
<td>-70.1</td>
</tr>
</tbody>
</table>

* Titrations of RPP21 (~200 μM) into RPP29 (~20 μM) were carried out in the standard ITC buffer with a varying amount of KCl at 55°C.

$^a$ The measured $\Delta H$ from this set of runs were quite different from the previous measurements (Table 3.1). Based on the thermodynamic parameters in Table E.1, we concluded that the $\Delta H$ in Table 3.1 were more trustworthy. However, both sets of experiments reported a similar $K_A$. Thus the slope of the linear fit to log $K_A$ as a function of log[KCl] would not be affected. The difference in $\Delta H$ observed may arise from the slight difference in buffer compositions during preparation.
Table E.3. Table of the thermodynamic parameters measured at different salt concentrations at 10°C. *

<table>
<thead>
<tr>
<th>[KCl] (mM)</th>
<th>Run</th>
<th>Stoichiometry</th>
<th>$K_A$</th>
<th>$ΔH$</th>
<th>$ΔS$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1</td>
<td>0.894 ± 0.004</td>
<td>0.847 ± 0.084</td>
<td>17.25 ± 0.12</td>
<td>92.6</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>1.02 ± 0.005</td>
<td>0.734 ± 0.085</td>
<td>17.54 ± 0.15</td>
<td>93.3</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>0.954 ± 0.003</td>
<td>1.18 ± 0.103</td>
<td>16.51 ± 0.09</td>
<td>90.6</td>
</tr>
<tr>
<td>50</td>
<td>1</td>
<td>1.02 ± 0.003</td>
<td>7.60 ± 1.5</td>
<td>17.94 ± 0.12</td>
<td>99.4</td>
</tr>
<tr>
<td>50</td>
<td>2</td>
<td>0.968 ± 0.002</td>
<td>10.1 ± 1.4</td>
<td>18.54 ± 0.08</td>
<td>102</td>
</tr>
<tr>
<td>50</td>
<td>3</td>
<td>0.988 ± 0.002</td>
<td>8.57 ± 1.3</td>
<td>15.67 ± 0.07</td>
<td>91.6</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>0.975 ± 0.003</td>
<td>17.3 ± 4.8</td>
<td>20.13 ± 0.14</td>
<td>109</td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>1.01 ± 0.002</td>
<td>23.2 ± 4.3</td>
<td>18.92 ± 0.08</td>
<td>105</td>
</tr>
<tr>
<td>100</td>
<td>3</td>
<td>0.993 ± 0.002</td>
<td>18.9 ± 4.6</td>
<td>18.41 ± 0.10</td>
<td>103</td>
</tr>
<tr>
<td>150</td>
<td>1</td>
<td>0.986 ± 0.003</td>
<td>24.6 ± 7.1</td>
<td>21.44 ± 0.14</td>
<td>114</td>
</tr>
<tr>
<td>150</td>
<td>2</td>
<td>0.966 ± 0.002</td>
<td>25.2 ± 6.7</td>
<td>18.52 ± 0.09</td>
<td>104</td>
</tr>
<tr>
<td>150</td>
<td>3</td>
<td>0.940 ± 0.002</td>
<td>22.4 ± 5.9</td>
<td>18.30 ± 0.11</td>
<td>103</td>
</tr>
</tbody>
</table>

*Titrations of RPP21 (~200 μM) into RPP29 (~20 μM) were carried out in the standard ITC buffer with a varying amount of KCl at 10°C.

Table E.4. Table of the thermodynamic parameters measured to obtain $ΔC_p$ in 150 mM KCl. *

<table>
<thead>
<tr>
<th>Temp (°C)</th>
<th>Run</th>
<th>Stoichiometry</th>
<th>$K_A$</th>
<th>$ΔH$</th>
<th>$ΔS$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1</td>
<td>0.986 ± 0.003</td>
<td>24.6 ± 7.1</td>
<td>21.44 ± 0.14</td>
<td>114</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>0.966 ± 0.002</td>
<td>25.2 ± 6.7</td>
<td>18.52 ± 0.09</td>
<td>104</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>0.940 ± 0.002</td>
<td>22.4 ± 5.9</td>
<td>18.30 ± 0.11</td>
<td>103</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>0.968 ± 0.006</td>
<td>29.3 ± 17</td>
<td>8.485 ± 0.119</td>
<td>67.7</td>
</tr>
<tr>
<td>35</td>
<td>1</td>
<td>0.967 ± 0.008</td>
<td>15.2 ± 8.5</td>
<td>-9.545 ± 0.166</td>
<td>6.46</td>
</tr>
<tr>
<td>45</td>
<td>1</td>
<td>0.976 ± 0.006</td>
<td>61.2 ± 5.1</td>
<td>-22.06 ± 0.37</td>
<td>-29.1</td>
</tr>
<tr>
<td>55</td>
<td>1</td>
<td>1.07 ± 0.005</td>
<td>7.71 ± 2.1</td>
<td>-32.07 ± 0.30</td>
<td>-61.6</td>
</tr>
<tr>
<td>55</td>
<td>2</td>
<td>1.01 ± 0.002</td>
<td>7.72 ± 1.2</td>
<td>-31.85 ± 0.16</td>
<td>-61.1</td>
</tr>
<tr>
<td>55</td>
<td>3</td>
<td>0.869 ± 0.003</td>
<td>7.26 ± 1.6</td>
<td>-34.80 ± 0.27</td>
<td>-70.1</td>
</tr>
</tbody>
</table>

*All experiments were carried out in the standard ITC buffer with 150 mM KCl. Titrations of RPP21 (~150 μM) into RPP29 (~15 μM) were performed at different temperatures ranging from 10°C to 55°C.
Table E.5. Table of the thermodynamic parameters measured in buffers with different ionization enthalpies at pH 6.7, 55°C. *

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Run</th>
<th>Stoichiometry</th>
<th>$K_A^{b}$</th>
<th>$\Delta H$</th>
<th>$\Delta S$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>($10^6 \text{ M}^{-1}$)</td>
<td>(kcal mol$^{-1}$)</td>
<td>(cal mol$^{-1}$ K$^{-1}$)</td>
</tr>
<tr>
<td>Cacodylate</td>
<td>1</td>
<td>1.06 ± 0.003</td>
<td>4.39 ± 0.24</td>
<td>-28.67 ± 0.13</td>
<td>-57.0</td>
</tr>
<tr>
<td>Cacodylate</td>
<td>2</td>
<td>1.04 ± 0.006</td>
<td>3.19 ± 0.27</td>
<td>-30.77 ± 0.25</td>
<td>-64.0</td>
</tr>
<tr>
<td>Cacodylate</td>
<td>3</td>
<td>0.980 ± 0.003</td>
<td>4.70 ± 0.29</td>
<td>-29.99 ± 0.16</td>
<td>-60.9</td>
</tr>
<tr>
<td>MOPS</td>
<td>1</td>
<td>1.02 ± 0.011</td>
<td>0.805 ± 0.056</td>
<td>-43.64 ± 0.63</td>
<td>-106</td>
</tr>
<tr>
<td>MOPS</td>
<td>2</td>
<td>1.04 ± 0.01</td>
<td>0.887 ± 0.058</td>
<td>-42.72 ± 0.58</td>
<td>-103</td>
</tr>
<tr>
<td>ACES</td>
<td>1</td>
<td>1.00 ± 0.006</td>
<td>2.22 ± 0.14</td>
<td>-48.39 ± 0.41</td>
<td>-118</td>
</tr>
</tbody>
</table>

* Titrations of RPP21 (~150 μM) into RPP29 (~15 μM) were carried out at 55°C except for experiments performed in cacodylate buffer (200 μM of RPP29 and 20 μM of RPP21). Each buffer contains 20 mM buffering reagent (cacodylate, MOPS or ACES), 10 mM KCl, 0.3 mM ZnCl$_2$, and 0.02% NaN$_3$. The pH of each buffer was adjusted to 6.7 at 55°C.
Figure. E.1. Representative isotherms of titrations of RPP29 into RPP21 over a temperature range from 10°C to 45°C in the standard ITC buffer with 10 mM KCl. (thermodynamic parameters shown in Table E.1)
Figure E.1 continued
Figure. E.2. Representative isotherms of titrations of RPP21 into RPP29 in the standard ITC buffer with different concentration of KCl at 55°C. (thermodynamic parameters shown in Table E.2)
Figure E.3. Representative isotherms of titrations of RPP21 into RPP29 in the standard ITC buffer with different concentration of KCl at 10°C. (thermodynamic parameters shown in Table E.3)
Figure. E.4. Representative isotherms of titrations of RPP21 into RPP29 in the standard ITC buffer with 150 mM KCl at temperatures ranging from 10°C to 55°C. (thermodynamic parameters shown in Table E.4)
Figure. E.5. Representative isotherms of titrations of RPP21 into RPP29 in buffers with different ionization enthalpies at pH 6.7, 55°C. (thermodynamic parameters shown in Table E.5)
BIBLIOGRAPHY


7  Smith D, Pace NR: Multiple magnesium ions in the ribonuclease P reaction mechanism. Biochemistry 1993, 32:5273-5281.


11 Sun L, Campbell FE, Zahler NH, Harris ME: Evidence that substrate-specific effects of C5 protein lead to uniformity in binding and catalysis by RNase P. EMBO J 2006, 25:3998-4007.

12 Sun L, Harris ME: Evidence that binding of C5 protein to P RNA enhances ribozyme catalysis by influencing active site metal ion affinity. RNA 2007, 13:1505-1515.

13 Kurz JC, Fierke CA: The affinity of magnesium binding sites in the Bacillus subtilis RNase P x pre-tRNA complex is enhanced by the protein subunit. Biochemistry 2002, 41:9545-9558.

14 Loria A, Pan T: Recognition of the T stem-loop of a pre-tRNA substrate by the ribozyme from Bacillus subtilis ribonuclease P. Biochemistry 1997, 36:6317-6325.


16 Wegscheid B, Hartmann RK: The precursor tRNA 3'-CCA interaction with Escherichia coli RNase P RNA is essential for catalysis by RNase P in vivo. RNA 2006, 12:2135-2148.


42 Perederina A, Esakova O, Quan C, Khanova E, Krasilnikov AS: Eukaryotic ribonucleases P/MRP: the crystal structure of the P3 domain. *EMBO J* 2010, ::.


85 Spolar RS, Record MTJ: Coupling of local folding to site-specific binding of proteins to DNA. Science (80-) 1994, 263:777-784.


93 Keller R: The Computer Aided Resonance Assignment Tutorial. 2004, :.


117 Fukada H, Takahashi K: Enthalpy and heat capacity changes for the proton dissociation of various buffer components in 0.1 M potassium chloride. Proteins 1998, 33:159-166.


