Structural insight of transcriptional terminator recognition by ProQ/FinO domain RNA chaperones

by

Hyeong Jin Kim

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry
University of Alberta

Abstract

The ProQ/FinO family is an emerging group of RNA binding proteins that are critical for sRNA-mediated gene regulation in proteobacteria. The ProQ/FinO domain of these proteins recognizes a transcriptional terminator, hairpin-tail structure, to regulate gene expression. The important role of the ProQ/FinO domain in RNA binding was well-studied, however, the molecular mechanism remained unclear due to a lack of structural information of ProQ/FinO proteins in complex with their target RNAs. To understand how the ProQ/FinO domain recognizes RNA in molecular detail, we chose to study RocC, a ProQ/FinO domain containing protein and its sRNA partner RocR.

The structure of the ProQ/FinO domain of RocC in complex with the RocR transcriptional terminator was determined using X-ray crystallography at 3.2 Å resolution. Two conserved structural motifs within RocC were identified that were the major contact points for RocR. One was an N-cap motif which recognized the A-form structure of one strand at the base of the RocR hairpin. The other was a positively charged pocket containing a β-turn-α-helix motif that recognized the last two 3' nucleotides of the 3' single-stranded tail.

Structure-guided mutagenesis demonstrated essential residues for RocC:RocR binding *in vitro* and transformation repression *in vivo*. As the structure revealed, mutations in the N-cap motif and the 3' binding pocket weakened binding and impaired biological function of RocC. Binding assays of RocC with a 10 nucleotide single-stranded RNA, which showed direct contact with RocC in the crystal structure, was tested using isothermal calorimetry (ITC). The dissociation constant (K_D) decreased 19-fold compared to the intrinsic terminator, suggesting that A-form helical structure is required for tight binding to restrain conformation of 3' side lower stem. The

substitution of the 3’ hydroxyl group to a phosphate completely disrupted binding with RocC. This result corroborated the 3’ hydroxyl group recognition by RocC observed in crystal structure.

An analysis of the conservation of the RocC RNA contact surfaces across the ProQ/FinO family suggests that the principles of RNA recognition from the RocC:RocR structure will hold for many members of this family.

Preface

The studies in this thesis are original works by Hyeong Jin Kim with collaborations and guidance as follows:

Works in Chapter 1 - 6 of this thesis were submitted and are under revision at *Nature Communications* (July 23, 2022).

In Chapter 3, most of the EMSA assays were done by Hyeong Jin Kim with help of Shiyun Peng, Jikun Su, Mazzen Black in the Glover lab.

In Chapter 4, protein:RNA purification and crystallization was done by Hyeong Jin Kim. Data collection and structure determination were done by Hyeong Jin Kim guided by Dr. Ross Edwards, Glover lab. The NMR study was performed by our collaborators in the Kreutz and Tollinger laboratories, University of Innsbruck. SAXS studies were a collaboration with Dr. Michal Hammel, SIBYLS Beamline Advanced Light Source Lawrence Berkeley National Laboratory.

In Chapter 5, site-mutagenesis for *in vitro* binding assay was performed by Dr. Rashmi Panigrahi (Glover lab), and the mutated RocC constructs were purified by Mazzen Black. FP experiments with mutated RocC were done by Hyeong Jin Kim, FP with ProQ and FinO were done by Mazzen Black. ITC was performed by Hyeong Jin Kim. Transformability assays were performed by our collaborator, Dr. Laetitia Attaiech, Charpentier lab, Université Claude Bernard Lyon.

In Appendix A, all works were done by Hyeong Jin Kim.

The work in Appendix B was published in *Structure*, 2019. Hyeong Jin Kim performed EMSA, SEC-SAXS and analyzed performed EMSA, SEC-SAXS data.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr. Mark Glover, for giving me an opportunity to participate in exciting projects and for his guidance throughout my PhD program. His patience, trust and support, and daily dose of morning coffee gave me the energy to move forward through my PhD program.

I want to give thanks to Dr. Rashmi Panigrahi for her unsparing help throughout my PhD program. Since the first year of my program, Rashmi patiently taught me all required techniques and gave me myriad scientific advice. Rashmi is a great mentor and friend. I will miss the Tim Horton coffee breaks we had.

My structural biology knowledge and skills can't be discussed without mentioning Dr. Ross Edwards. From the crystal mounting to structure determination, he taught me everything and shared his years of experience and skill. Especially, I will never forget the day Ross called to show me the electron density of RNA in our crystal structure.

All Glover lab members also gave me unforgettable memories, the cross-country ski trip to Yoho national park, the overnight crystal shooting with donair pizza, also our yearly Christmas party. I will think about the precious memories we shared together Ross, Jun, Rashmi, Rabih, Cameron and Mazzen.

I appreciate my committee members for invaluable advice, Dr. Andrew MacMillan, Dr. Tracy Raivio, Dr. Charles Holmes and internal examiner for Dr. Richard Fahlman. I also thank Dr. Gisela Storz, who was happy to be my external examiner for my defense.

I also appreciate for my family, Jae-ho Kim (father), Young-hee Lee (mother), and Su-jin Kim (younger sister) who always supported my graduate study across the Pacific Ocean.

Last but the most important, my fiancé, Khanh C. N. Trần, thank you for being my side and supporting me through my PhD program.

Table of Contents

Chapter 2. Materials and methods **34**

List of Tables

List of Figures

List of Abbreviations

ADAR - adenosine deaminase RNA-specific binding protein

ALS – Advanced Light Source

ApU – adenine uracil monophosphate

A-rich – adenine-rich

ASPP – apoptosis-stimulating protein of p53

B3H – bacterial three-hybrid

Cas- CRISPR-associated

CLS – Canadian Light Source

CRISPR - clustered regularly interspaced short palindromic DNA repeats

crRNA – crispr RNA

cryo-EM – cryogenic electron microscopy

CUGBP – CUG-binding protein

DEPC – diethyl Pyrocarbonate

DLS – dynamic light scattering

DNA – deoxyribonucleic acid

dsRBD – double-stranded RNA binding domain

DTT – 1,4-Dithiothreitol

EFA – evolving factor analysis

EMSA – electrophoretic mobility shift assay

FAM - fluorescein amidites

FP - fluorescence polarization

FRET – fluorescence resonance energy transfer

GpC – guanine cytosine monophosphate

GST - glutathione S-transferase

HDX – hydrogen deuterium exchange

IPTG – Isopropyl ß-D-1-thiogalactopyranoside

ITC – isothermal calorimetry

LB – luria broth

MDP – 10-Methacryloyloxydecyl dihydrogen phosphate

miRNA – micro RNA

MR – molecular replacement

mRNA – messenger RNA

NCS – non-crystallographic symmetry

NLS – nuclear localization signal

NMR - nuclear magnetic resonance

NOESY – nuclear overhauser effect spectroscopy

PCR – polymerase chain reaction

PDB – protein data bank

PKR - protein kinase R

PP-1c – protein phosphatase 1 catalytic subunit

RBP – RNA binding protein

RBS – ribosome-binding site

RMSD – root-mean-square deviation

RNA – ribonucleic acid

RNase - ribonuclease

RRM – RNA recognition motif

rRNA – ribosomal RNA

SAM - S-adenosylmethionine

SD sequence – Shine-Dalgarno sequence

SDS-PAGE - sodium dodecyl sulfate–polyacrylamide gel electrophoresis

SEC-MALS – size-exclusion chromatography with multi-angle light scattering

SEC-SAXS - size-exclusion chromatography with small-angle X-ray scattering

SELEX – systematic evolution of ligands by exponential enrichment

sgRNA - single-guided RNA

siRNA – small interfering RNA

SL – stem-loop

sRNA – small RNA

ssRNA – single-stranded RNA

SVD – singular value decomposition

T4PNK – T4 polynucleotide kinase

TCEP – Tris(2-carboxyethyl)phosphine

TMP – thiamine monophosphate

TPP - thiamine pyrophosphate

tracrRNA – trans-activating crRNA

tRNA – transfer RNA

TSA – thermal shift assay

TSA – thermal shift assay

UMP – uridine monophosphate

UTR – untranslated region

UV-CLIP - ultraviolet cross-linking immunoprecipitation

WT – wild type

XIST – X inactivate specific transcript

Chapter 1.

Introduction

1.1. RNA is central to the regulation of gene expression in all kingdoms of life

Three decades ago, scientists predicted that the protein-coding genome would be "only" 3% of the human genome (It turned out to be approximately 1%), and 97% of the remaining non-coding DNA was termed junk DNA (Nowak, 1994). However, a large body of research over the last decades have revealed increasingly diverse roles for non-translated RNAs in all kingdoms of life. For example, the XIST (X inactive specific transcript) gene was reported as an essential regulator for inactivating the X chromosome in mouse and humans (Brockdorff et al., 1992). The XIST gene is approximately 15k base-pair in size and lacks an open reading frame. XIST was shown to regulate the X chromosome directly, indicating its inhibitory function is rooted in the XIST (RNA). Subsequent work has revealed different classes of regulatory RNAs in eukaryotes, such as miRNA and siRNA, which expand the knowledge of regulatory RNAs (Morris & Mattick, 2014).

The fraction of non-coding DNA in prokaryotes is smaller than in eukaryotes, but it is still significant. An analyses of 39 bacteria and 11 archaea revealed that the average portion of non-coding DNA was 18%, except for *Mycobacterium leprae* (50%) and *Rickettsia prowazekii* (24%) (Rogozin et al., 2002). The growing discoveries of RNA's regulatory functions changed the concept of non-coding RNA from junk DNA to RNA regulators in gene expression (Inouye & Delihast, 1988; Karen Montzka Wassarman et al., 1999). An early example of a regulatory RNA system in prokaryotes was RNAI and RNAII in the ColE1 plasmid (Stougaard et al., 1981; Tomizawa et al., 1981). RNAI and RNAII are *cis*-acting RNAs that are encoded in opposite strands. RNA II is processed by RNase H, and this digested RNAII serves as a primer for plasmid. This activity of RNA II is however repressed by interactions with its antisense RNA partner, RNAI. Duplex formation is initiated by kissing interactions between complementary loops from the two RNAs, as revealed by solution state NMR. This loop-loop interaction between RNAI and RNAII requires the protein Rop (also referred to as Rom) to enhance the duplex formation of non-kissing parts of the RNAs and to stabilize these interactions (Tomizawa & Som, 1984). Subsequent discoveries have revealed other classes of regulatory RNAs in addition to antisense RNAs. These include sRNA regulators(Gottesman & Storz, 2011; Inouye & Delihast, 1988), regulatory elements in mRNA 5' UTRs (often called riboswitches) (Garst et al., 2011) CRISPR mediated adaptive immunity (Ishino et al., 2018; Jiang & Doudna, 2017).

1.1.1. Advantages of RNA-based gene regulation

A large portion of gene regulation in bacteria is dependent on RNA regulators compared to eukaryotes. For example, the thiamine pyrophosphate (TPP) riboswitch is the only reported riboswitch system in eukaryotes to date (Winkler et al., 2002) on the other hand, riboswitches in *Bacillus subtilis* control ~4% of genes (Mandal et al., 2003). Transcriptomics has revealed that the major bacterial RNA chaperone Hfq, together with its interacting sRNAs, influences more than 20% of genes in *Escherichia coli* and *Salmonella enterica* (Chao et al., 2012; Tree et al., 2014).

RNA regulators enable bacteria to respond quickly to changing environments. DNA stores genetic information which must be first transcribed into mRNA and subsequently translated into protein. RNA-based methods allows regulation at different stages along this information flow, potentially at the transcriptional, post-transcriptional (translational), or post-translational levels. RNA regulation is often at the post-transcriptional level, targeting the mRNA to control its translation. This form of regulation has been described as an efficient mechanism which conserves valuable cellular resources (Beisel & Storz, 2010; Liang et al., 2011; Waters & Storz, 2009). Other potential advantages of RNA regulation have been reviewed (Gottesman & Storz, 2011; Narberhaus & Vogel, 2009; Waters & Storz, 2009). In the next sections I will discuss three different classes of regulatory RNAs that play major roles in the regulation of bacterial gene expression: the riboswitch, CRISPR, and sRNA-mediated gene regulation.

1.1.2. Riboswitch

Some mRNAs contain regulatory elements that can regulate their expression by controlling their transcription and translation in response to changing environments. These regulatory elements are called riboswitches, and are often located in mRNA 5’ UTR. Riboswitches comprise two modular domains: an aptamer domain and an expression platform (Figure 1-1A). Ligand recognition through the aptamer domain triggers the rearrangement of the expression platform to an alternative secondary structure. Aptamer ligands can be diverse, including small chemicals, coenzymes, metal ions, S-adenosylmethionine (SAM), and even tRNA (Sherwood & Henkin, 2016). Riboswitches are an important mode of regulation in bacteria. For example, approximately 4% of *B. subtilis* genes are under the control of riboswitches (Mandal et al., 2003). The TPP

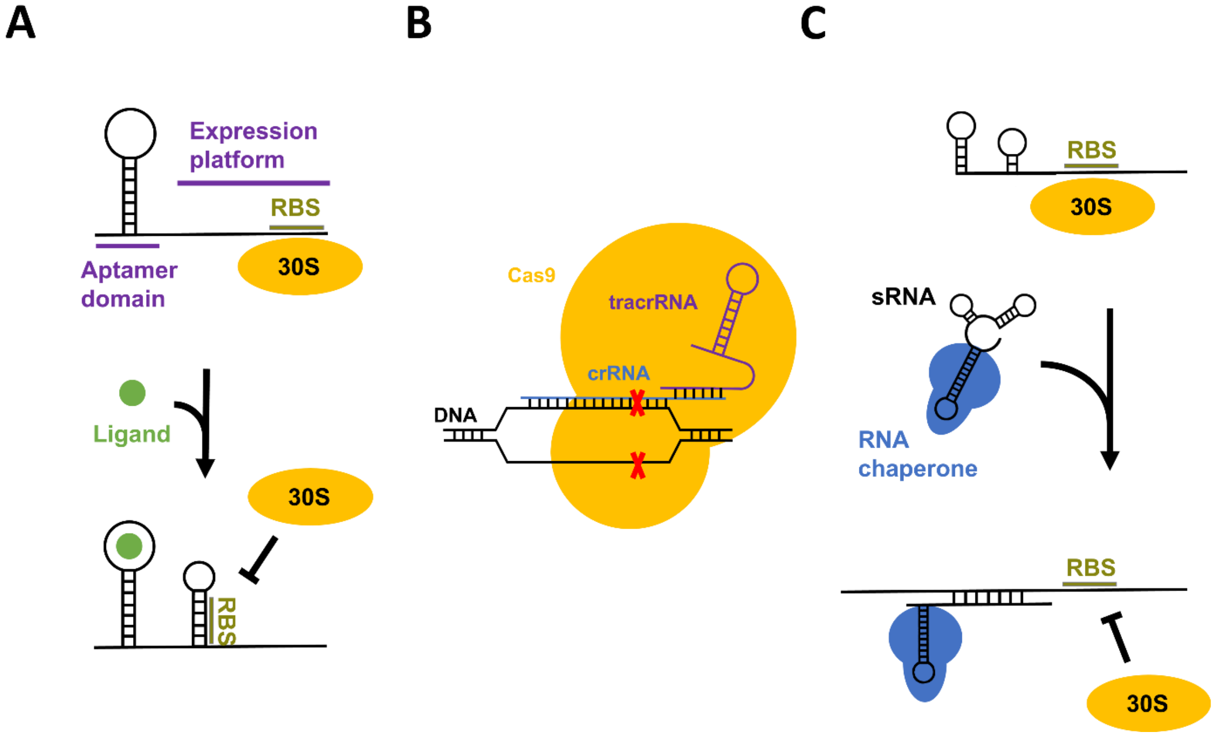


Figure 1-1. Regulatory RNAs in bacteria.

(A) Riboswitch. (B) CRISPR-Cas9. (C) sRNA-mediated gene regulation.

riboswitch is one of the earliest reported and well-studied systems (Mironov et al., 2002; Winkler et al., 2002). This well-conserved system is found in bacteria, archaea, and even fungi (Griffiths-Jones et al., 2005; Rodionov et al., 2002; Sudarsan et al., 2003). In *E. coli*, two distinct TPP riboswitches exist, *thiM* and *thiC*, which are involved in the thiamine biosynthesis. Both are located in the 5' UTR region and regulate translation initiation by interacting with TPP through the *thi* box (Ontiveros-Palacios et al., 2008; Winkler et al., 2002). Aptamer binding causes the steric hindrance of the AUG start codon, thereby repressing translation initiation. Both *thiM* and *thiC*, bind their thiamine-based ligands with high affinity (Winkler et al., 2002). For example, *thiM* bound TPP ~10-fold and 100-fold more tightly than TMP and thiamine, respectively. *thiC* bound TPP ~10-fold tighter than *thiM,* and the selectivity for TPP compared to TMP and thiamine was greater than 1,000-fold (Winkler et al., 2002). Riboswitches are also involved in gene expression at the transcriptional level. The *thiM* TPP riboswitch is located in the 5' UTR of the *thi*MD operon, which encodes essential proteins for thiamine biosynthesis. Binding of TPP to the *thiM* riboswitch triggers the formation of a Rho-independent terminator that leads to premature transcription termination (Bastet et al., 2017).

1.1.3. CRISPR-RNA

Clustered regularly interspaced short palindromic DNA repeats (CRISPR) were discovered in *E. coli* more than 30 years ago. As the name implies, repeated sequences are separated by 20-60 base-paired spacer elements. Sequence analysis revealed that similar repeats were found in various bacteria and archaea, meanwhile, the spacer DNA was highly similar to existing DNA from bacteriophage and conjugative plasmids (Mojica et al., 2005). Bolotin and colleagues discovered the inverse correlation between phage infection sensitivity and the number of spacers from the particular phage (Bolotin et al., 2005). Focusing on nuclease genes (*cas*) adjacent to CRISPR elements, they suggested that the acquired spacer sequence might target exogenous donor DNA and cleave DNA in association with the Cas5 endonuclease (COG3513, Cas5 is renamed to Cas9). The correlation between phage resistance and the spacer was directly tested by inserting the phage genome into the spacer region, resulting in resistance to phage infection (Barrangou et al., 2007). The deletion of the inserted gene made the bacteria vulnerable to phage infection. Strikingly, deletion of adjacent *cas*5 (*cas*9) gene impaired the acquired immunity in the presence of the inserted spacer DNA from the donor. CRISPR DNA repeats transcribed into pre-crRNA and tracrRNA, small noncoding RNA, binds through the repeated sequence in pre-crRNA. Formation of the pre-crRNA:tracrRNA complex leads to crRNA maturation by RNase III digestion (Deltcheva et al., 2011). tracrRNA and crRNA forms the gRNA, or guide RNA, complex, which then recruits the Cas protein. The single-stranded crRNA guides this complex to specific complementary DNA sequence (Figure 1-1B). The CRISPR and Cas protein work together to achieve immunity against the phage. The CRISPR-Cas mechanism of targeted DNA cleavage was probed by biochemical experiments (Jinek et al., 2012). Point mutations in the HNH and RuvC-like domains of the CRISPR nuclease demonstrated that these individual domains digest the complementary and noncomplementary strands, respectively. The two RNAs, tracrRNA and crRNA, can be connected through a short linker named single-guided RNA (sgRNA) without disturbing nuclease activity. The understanding of the CRISPR system expanded the physiological roles of regulatory RNA in prokaryotes and provided a novel platform for gene editing (Jinek et al., 2012).

1.1.4. sRNA-mediated gene regulation

Bacterial sRNAs (small RNAs) are short transcripts ranging from 50 to 300 nucleotides in size (Beisel & Storz, 2010; Gottesman & Storz, 2011). sRNAs generally modulate gene expression through interactions with target mRNAs, often with the assistance of accessory RNA chaperone proteins.

sRNA-mediated gene regulation is achieved mainly by base-pairing with mRNA (Figure 1-1C) (Beisel & Storz, 2010; Gottesman & Storz, 2011). sRNA often downregulates target gene expression by interacting through the mRNA 5' UTR (Arthur et al., 2003; Baker et al., 2007; Smirnov et al., 2017), near or at the ribosome binding site (RBS) to block its interaction with the 3' end of the 16s rRNA. Sometimes, sRNA base-pairing with the mRNA affects stability to promote RNase-induced degradation of the duplex. It is less common than in the previous case, but sRNAs can also upregulate the target gene (Papenfort & Vanderpool, 2015). In this case, the RBS is sequestered in a secondary structure, but becomes exposed upon interaction of the sRNA with the mRNA and subsequent conformational rearrangement.

sRNA-mediated gene regulation can be divided into *cis*-acting and *trans*-acting depending on where the sRNA was derived (Beisel & Storz, 2010; Gottesman & Storz, 2011). *Cis*-acting sRNAs (antisense RNA) are encoded in the strand complementary to the target strand. This perfect complementarity provides precise regulation toward a single target RNA. On the other hand, the genomic locus of *trans*-acting sRNA is not dependent on the locus of the target gene. Trans-acting sRNAs usually only pair with their targets through more limited base pairing. Due to this imperfect complementarity, one sRNA can target multiple genes (Attaiech et al., 2016) or multiple sRNAs can regulate a single mRNA (De Lay & Gottesman, 2012). Due to the highly structured nature of many of these RNAs, they often require specialized RNA binding proteins, called RNA chaperones, to remodel their structure to achieve their biological roles.

1.2. History of RNA structural biology: from the fiber diffraction to high-resolution studies

The knowledge of RNA structures in textbooks results from decades of efforts and passion from structural biologists. In 1956, 3 years after the double-helical structure of DNA, Rich and Davies obtained fiber X-ray diffraction of the poly A:U RNA duplex, showing the features of the double helix but different from the patterns obtained from double helical DNA. The first short double-helical RNA structure from dinucleotides of ApU and GpC at atomic resolution in 1973 visualized detailed features hypothesized from the earlier fiber diffraction. In 1965, Holley and colleagues successfully purified transfer RNA, tRNA, from yeast (Holley et al., 1965). RNase digestion of tRNA was used to propose conserved secondary structure models, all of which contained exposed D loop, P loop, and anticodon loops with a 3' overhang. In 1974, one of the three was predicted to be the now familiar cloverleaf structure which turned out to be conserved in tRNAs. Kim and colleagues determined the first crystal structure of a tRNA, yeast tRNA^{Phe} in the 1970s (S. H. Kim et al., 1974). This ~ 80 nucleotides tRNA structure determination provided insights on RNA folding. The tRNA structure was significantly folded in 3D, and revealed tertiary interactions such as coaxial stacking, loop-loop interaction, U-turn, etc (Figure 1-2A). The D-loop (7-11 nucleotides) and T-loop (5 nucleotides with consensus sequence, U(G/U)NR(A/U)) are connected to the anticodon hairpin, containing a U-turn (Figure 1-2B-D). The exposed D-loop and T-loop interact through the interdigitated stacking of bases and trans base-pairing involving both Watson-Crick and Hoogsteen interactions (Figure 1-2E, F). More RNA structures were determined through the 1990s, such as the hammerhead ribozyme and the group I intron. The accumulation of RNA structures confirmed various secondary structures predicted through phylogenetic comparison. In addition, these structures visualized the formations of tertiary structures involving unusual contacts through internal loops and hairpin loops. For example, the P4-P6 domain of the 160 nucleotide group I intron provided the three-dimensional structures of secondary structural motifs, such as the A-rich bulge, tetraloop, internal loop, and hairpin loops, and tertiary structural motifs, such as coaxial stacking, the ribose zipper, and tetraloop-receptor (Cate et al., 1996). The landmark determination of the crystal structures of the 30S and 50S ribosomal subunits, which consist of ~1500 nucleotides and ~3000 nucleotides, respectively, greatly expanded the RNA structure database in the 2000s (Ban et al., 2000; Wimberly et al., 2000).

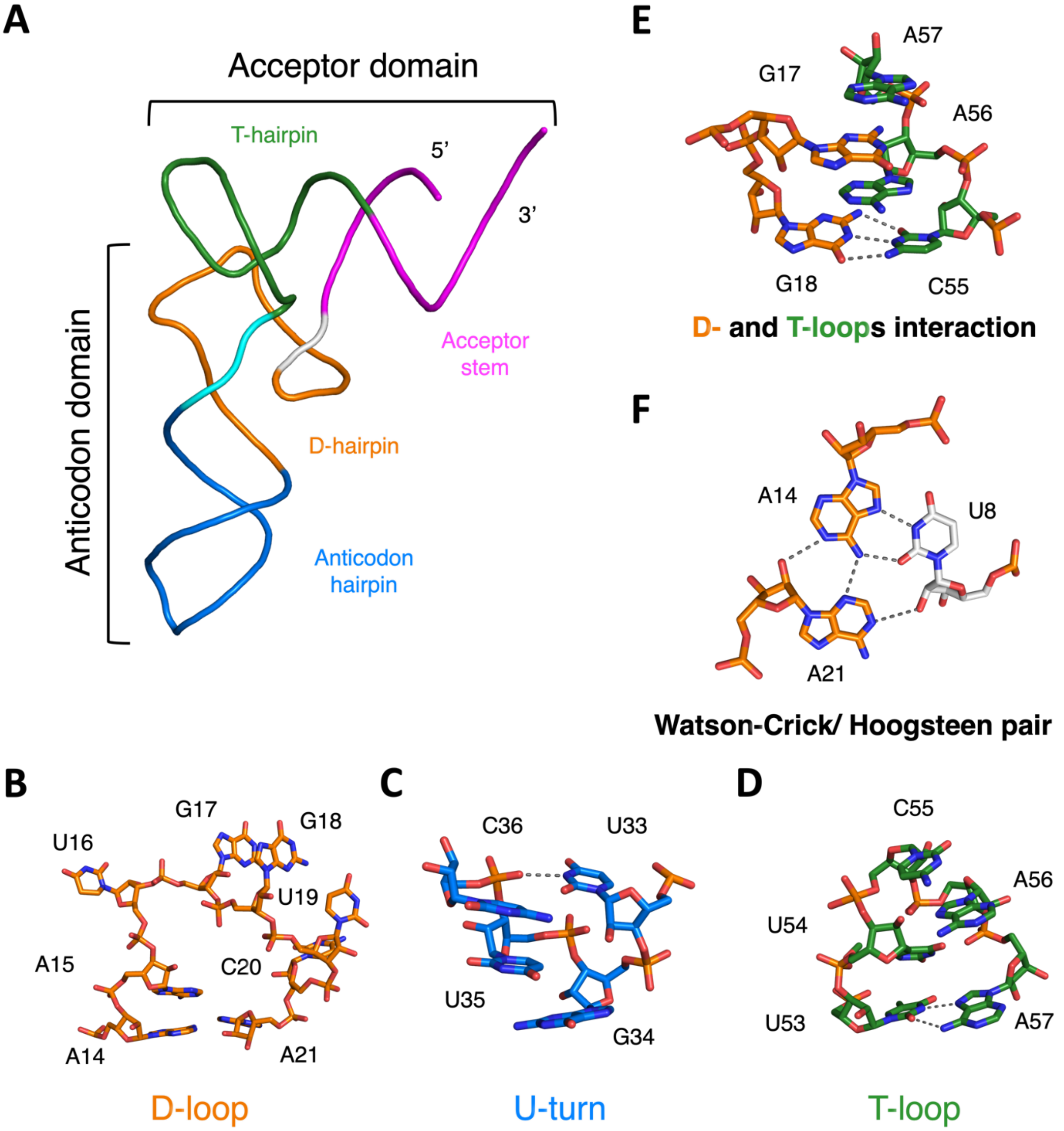


Figure 1-2. Structure of transfer RNA.

(A) Tertiary structure of tRNA$^{\text{Asp}}$. (B) Loop of D arm. (C) U-turn in anticodon hairpin. C36 phosphate oxygen and U33 base form a hydrogen bond to stabilize the sharp U-turn. G35, U35, and C36 are the anticodon triplet. (D) Loop of T arm. N and N+4 residues form A-U base pairing. (E) Tertiary interactions between the D-loop and T-loop. Stacking interactions between A57-G17-A56 and Watson-Crick base pairing between G18 and C55. (F) Watson-Crick/ Hoogsteen pairs between A 14 and U9, forming additional hydrogen bonds with the third base.

The understanding of RNA structural motifs promoted the understanding of protein:RNA interactions which is critical to understand their biological roles. In 1991, a few years earlier than structure of hammerhead ribozyme was revealed, the structure of tRNA^{Gln} in complex with glutaminyl-tRNA synthetase revealed the first structure of a protein bound to an RNA (Rould et al., 1991). The structure showed that the tRNA synthetase recognized the tRNA secondary structure, such as loop and stem, and single stranded regions. This was the first of many tRNA synthetase-tRNA complex structures, which revealed diverse ways in which these enzymes specifically recognized their cognate tRNAs. A few years later, the first structure of a non-tRNA RNA bound to a protein was determined; that of the spliceosomal U1A RNA recognition module (RRM) bound its U1A target RNA hairpin. The structure revealed the importance of conserved aromatic residues in the RRM that stack with unpaired bases in the hairpin loop (see further discussion of RRM domains below) (Chris Oubridge et al., 1994). Accumulation of structural data from complex ribonucleoprotein assemblies, such as ribosomes and spliceosomes in the early 2000s (Ban et al., 2000; Wimberly et al., 2000) to recent CRISPR-CAS9 structures, has expanded our understanding of how proteins interact with RNAs to perform their biological roles.

1.3. RNA binding proteins recognize target RNA through structure and sequence

RNA binding proteins (RBP) interact with RNA to perform gene regulation and diverse cellular roles. RBPs bind specifically to target RNAs by recognizing unique structural features or sequences. In general, eukaryotic proteins possess modular domain organizations and many RBPs contain multiple RNA binding motifs/domains which work together to recognize RNAs (Cléry & Allain, 2012). This modular architecture can be beneficial to increase specificity and binding affinity (Dong et al., 2004). Meanwhile, RBPs in prokaryotes tend to have a single or few RNA binding motifs/domains (Holmqvist & Vogel, 2018). I will discuss a few RBP domains and motifs, focusing on RNA target recognition from structural perspectives.

1.3.1. Double-stranded RNA binding domain

The double-stranded RNA binding domain (dsRBD) is a RNA recognition domain widely found in all kingdoms of life. The domain consists of approximately 70 amino acids and adopts an

$\alpha\beta_3\alpha$ fold. dsRBDs are often a part of RNA processing proteins, such as the RNase III family, ADAR (Adenosine deaminase RNA-specific binding protein), and PKR (protein kinase R), helping these proteins to specifically interact with double-stranded RNA targets (Gan et al., 2006).

Figure 1-3A shows the structure of the dsRBD from *Aquifex aeolicus* RNase III (Gan et al., 2006). The structure reveals three distinct protein/RNA interfaces. Two minor grooves are recognized by the $\alpha1$ helix and loop between $\beta1$ and $\beta2$ strands while a portion of the major groove is recognized by the N-terminus of the $\alpha2$ helix. Multiple hydrogen bonding interactions are made between amino acid residues and 2'-hydroxyl groups of the ribose sugars exposed in the wide A-form minor groove. There are also contacts to the bases via the minor groove as well as electrostatic recognition of the backbone phosphate groups (figure 1-3B, C, D).

1.3.2. RNA recognition motif, a single-stranded RNA recognition motif

Among the most abundant of all RNA binding domains is the RNA recognition motif (RRM), which is generally involved in the recognition of single-stranded RNA. RRMs are more prevalent in eukaryotes than in other kingdom. 497 proteins in humans were reported that contain at least one RRM, which is approximately 2% of human genes (Maris et al., 2005). More than two RRMs are often found in a protein, and multiple RRMs provide higher specificity and binding affinity. For example, CUG-binding protein 1 (CUGBP1) contains tandem RRM1 and RRM2 which together recognize their RNA targets with much higher affinity than the isolated domains (Teplova et al., 2010).

Canonical RRMs are approximately 90 amino acids in size and generally consist of two α-helixes and four β-strands. RRMs typically adopt a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology (Figure 1-4A). The RRM is also defined by two sequence motifs called RNP1 (on $\beta3$) and RNP2 (on $\beta1$) (Maris et al., 2005). These motifs are positioned in the center of the β-sheet and play an essential role in RNA interactions, via conserved aromatic residues that stack with the bases of the RNA target (Figure 1-4)(Cléry et al., 2008; Maris et al., 2005; Muto & Yokoyama, 2012). RRM1 from poly(A)-binding protein in complex with poly(A) tail is an example in which three conserved tyrosine residues stack with the RNA bases (Figure 1-4B).

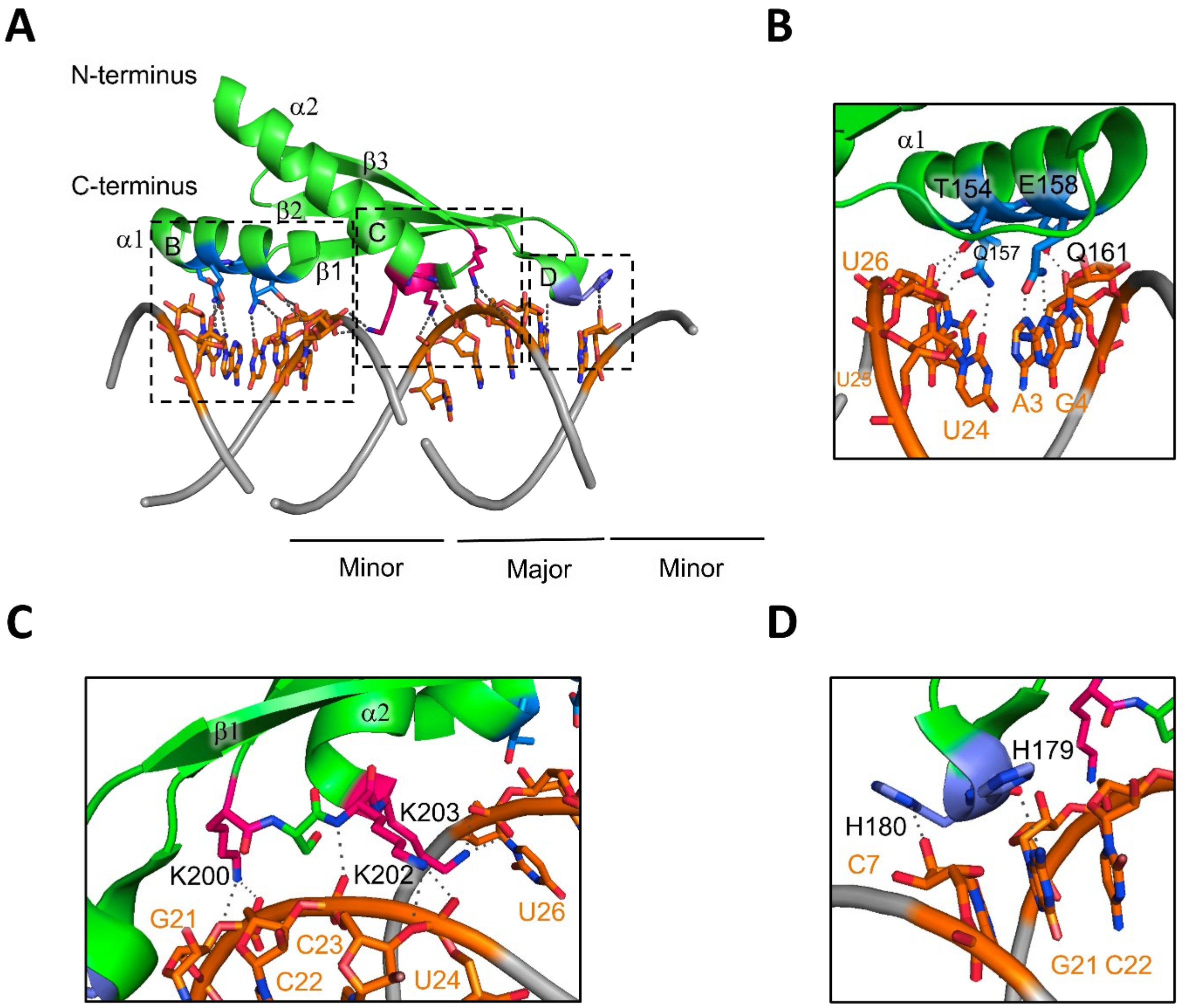


Figure 1-3. Structure of dsRNA in complex with dsRBD from *A. aeolicus* (PDB ID: 2EZ6).

(A) dsRBD interacts with dsRNA through three distinct surfaces. (B), (C), and (D) Zoomed in views of specific interactions.

1.3.3. La protein recognizes 3' poly(U) tails with specific recognition of the terminal 3' nucleotide

La protein is an RNA binding protein with roles in both the nucleus and cytoplasm of metazoans (Bayfield et al., 2021; Maraia & Bayfield, 2006). In the nucleus, the La protein binds to nascent Pol III transcripts to protect the RNA from ribonuclease, while in the cytoplasm La protein regulates translation. The canonical La motif is connected to an RNA recognition domain

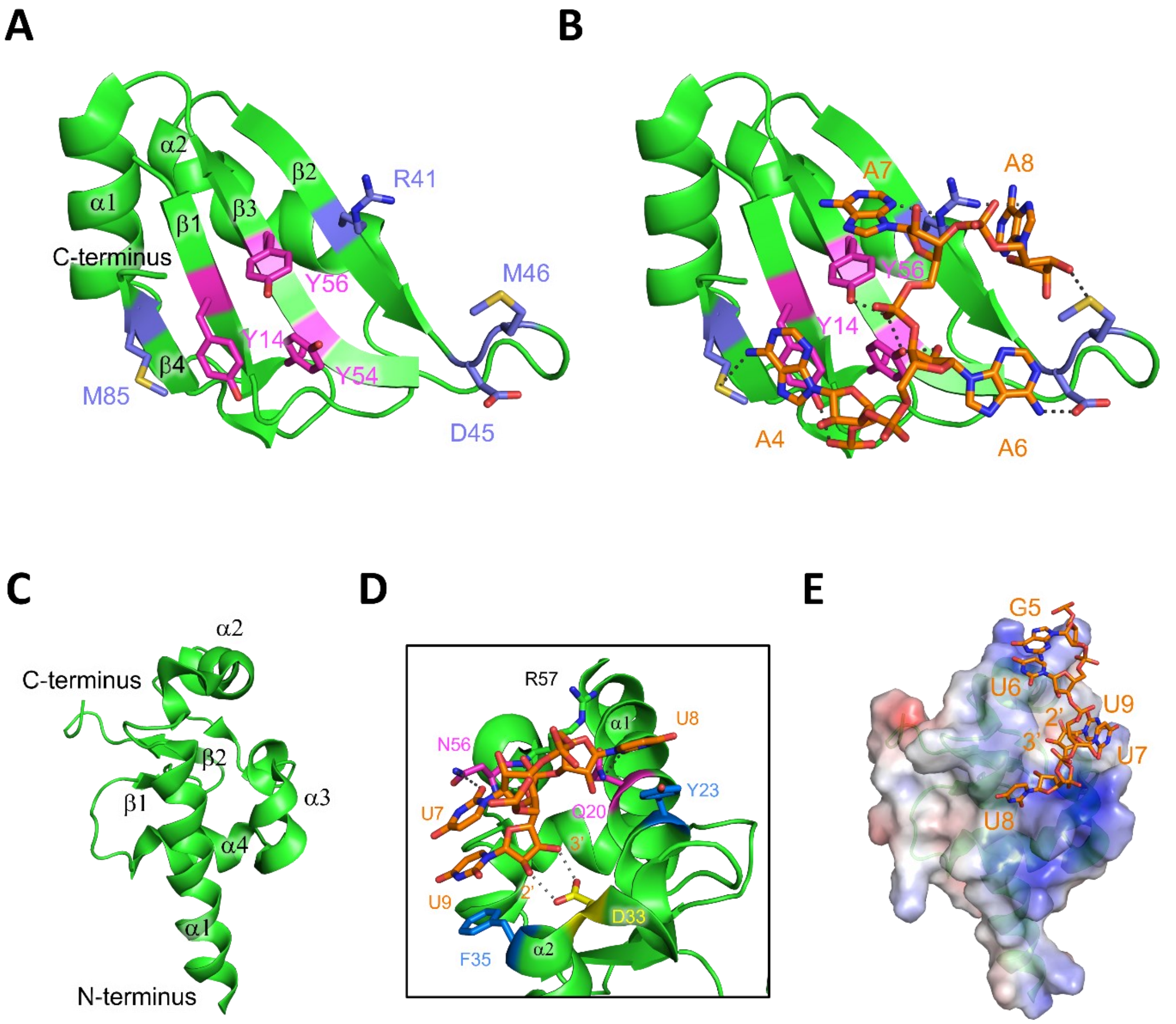


Figure 1-4. Structure of RRM and La motifs in complex with target RNAs.

(A) RRM structure (PDB ID: 1CVJ). Conserved aromatic rings are coloured magenta. Base recognition residues are coloured purple. (B) Structure of RRM in complex with a poly(A) tract. (C) La motif from human poly(A)-binding protein (PDB ID: 1ZH5) (D) La motif in complex with UUU-3'OH. Conserved aromatic rings are coloured with blue. Base recognition residues are coloured magenta. (E) Electrostatic potential map of La motif with poly(U) tract.

(RRM) to form a La module. In some organisms like yeast, La has a flexible C-terminus containing an NLS (nuclear localization signal). La proteins from higher eukaryotes have an additional RRM after the La module (Bayfield et al., 2021). La proteins recognize nascent RNA polymerase III transcripts through their terminal UUU-OH 3' motif to protect them against 3'-5' exoribonuclease digestion.

The La motif is comprised of ~90 amino acids (Figure 1-4C). The Human La motif has two highly conserved tyrosine residues that stack with the terminal uridine bases, as well as additional interactions from residues in the La motif and the RRM (Teplova et al., 2006). The recognition of the terminal 3' ribose is by Asp33, which forms hydrogen bonds through the 2' and 3' hydroxyl groups and provides direct recognition of the RNA chain terminus, and is critical for high-affinity recognition of RNA (Figure 1-4D) (Uchikawa et al., 2015). The decrease in La binding affinity by replacing the 3' hydroxyl group with phosphate further the demonstrates the importance of 3' hydroxyl recognition the (Maraia & Intine, 2001). Although the crystal structure showed the La motif contributes as a primary binding surface for RNA, biochemical assays demonstrated that the La module, composed of a La motif fused to RRM1, is an indispensable RNA binding unit. The binding affinity of full-length human La protein and La module for a 9 nucleotide poly(U) tract revealed that the La module alone binds 5-fold tighter than full-length La protein, indicating the La module is the main RNA binding surface (Ohndorf et al., 2001). Dong and colleagues tested the binding affinity of the individual motifs in the *Trypanosoma brucei* La protein (Dong et al., 2004). *T. brucei* full-length La protein and La modules showed similar binding affinities. However, both the La motif and RRM alone failed to bind RNA. Intriguingly, incubating the isolated La motif and RRM together with RNA still failed to show interactions with RNA, suggesting the La motif linked with RRM cooperatively binds to RNA.

1.4. RNA chaperones – remodellers of RNA structure

RNA is a structured and versatile molecule that performs a myriad of biological roles through its dynamic nature. Due to kinetic traps in RNA folding, alternative conformations can be achieved through the action of protein cofactors called RNA chaperones. For example, the *E. coli* sRNA RyhB is dependent on the RNA chaperone Hfq (Geissmann & Touati, 2004). RyhB is a non-coding small RNA comprising 90 nucleotides that folds into three stem-loops. RyhB has a complementary sequence with *sodB* mRNA, which encodes iron superoxide dismutase. The *sodB* mRNA forms two stem-loops at the 5' UTR, and the start codon is hidden in a bulge of the hairpin. RNase footprinting revealed that the conformation of the *sodB* mRNA hairpin, including the start codon, is rearranged by Hfq. This alternative form exposes the hidden complementary sequences and allows them to hybridize with RyhB. This duplex is subsequently digested by RNase E (Geissmann & Touati, 2004). Hfq serves as a major RNA chaperone that facilitates base-pair destabilization and the promotion of RNA annealing without the consumption of ATP (Herschlag, 1995; Rajkowitsch et al., 2007). In the following sections, I will summarize three dominant RNA chaperones in bacteria; Hfq, CsrA, and ProQ.

1.4.1. Hfq, a prominent RNA chaperone in bacteria

Hfq was discovered as a host factor for replicating bacteriophage QB RNA in *E. coli* (Franze De Fernandez et al., 1968). Hfq is the largest posttranscriptional regulator, widely spread throughout proteobacterial species. Hfq facilitates extensive sRNA-mediated gene regulation (Kavita et al., 2018; Vogel & Luisi, 2011). Despite its prevalence across bacteria, the global influence of Hfq on gene regulation systems are quite different in gram-negative and gram-positive bacteria. A microarray analysis in the gram negative bacteria *Brucella melitensis* demonstrated the global influence of Hfq in the response to stress, regulation of metabolism, and virulence (Cui et al., 2013). Some studies showed that Hfq and its associated sRNAs regulate more than 20% of genes in *E. coli* and *S. enterica* (Chao et al., 2012; Tree et al., 2014). Meanwhile, the role of Hfq in gram-positive bacteria such as *B. subtilis* is limited compared to in gram-negative bacteria. Transcriptome changes were compared in the presence and absence of Hfq (Hämmerle et al., 2014), revealing more limited impacts, such as on aerobic and anaerobic respiration control by ResD-ResE (Härtig & Jahn, 2012) and on the ComK regulon which controls competence development (Lopez et al., 2009). The impacts in sRNA abundance were limited in that only 6 out of ~100 known sRNAs were influenced by lack of Hfq (Hämmerle et al., 2014).

1.4.1.1. Hfq regulates gene expression through multiple RNA binding sites

Hfq is a small protein consisting of around 70~100 amino acids (Santiago-Frangos et al., 2017). The protomer adopts an Sm-like (Lsm) fold, consisting of an α-helix followed by five antiparallel β-strands with an the intrinsically disordered C-terminal tail of variable length (Figure 1-5A) (Santiago-Frangos et al., 2017; Stanek et al., 2017). Similar to other Lsm and Sm domain proteins (Mura et al., 2013), Hfq forms a toroidal homo-hexameric ring, mediated by extensive interaction between β4 and β5 (Figure 1-5A). The hexameric ring contains four distinct RNA binding surfaces (Figure 1-5B). The flat surface of the overall doughnut shape largely consists of the α-helix and is named the proximal face. The opposite side of the flat surface is named the distal face. The side of the ring is called the rim or lateral face.

1.4.1.1.1. The Hfq proximal face recognizes poly(U) RNAs

Poly(U) tracts are often found at the 3' ends of sRNAs as part of rho-independent transcriptional termination structures. The recognition of poly(U) tracts by Hfq is well defined (Ishikawa et al., 2012; Otaka et al., 2011a; Sauer & Weichenrieder, 2011). The structure of *E. coli* Hfq showed how the proximal face recognizes U-rich sequences (Figure 1-5C) (W. Wang et al., 2013). Phe42 intercalates between the bases in a stacking interaction. Gln8, Gln41 (Asn in *Listeria monocytogenes*), and Lys56 directly contact and recognize uridine bases through hydrogen bonding interactions. The 3' hydroxyl group at the terminus of the 3' tail is directly recognized by His57 in *S. enterica* Hfq and *E. coli* Hfq (Dimastrogiovanni et al., 2014; Sauer & Weichenrieder, 2011) in gram-negative bacteria and *L. monocytogenes* Hfq in gram-positive bacteria (Kovach et al., 2014). The importance of this interaction was demonstrated by replacing the 3' hydroxyl group with propyl-phosphate of U_6 (Sauer & Weichenrieder, 2011). The replacement of the 3' hydroxyl group showed a dramatic binding decrease, indicating the 3' hydroxyl group recognition by *S. enterica* Hfq. Schumacher and colleagues reported an initial Hfq structure from *S. aureus* bound to a short RNA (5'-AUUUUUG-3'). In this structure, the poly(U) tract encircles the proximal face. Individual bases are intercalated between Tyr42 (similar to the interactions of Phe42 in *E. coli* Hfq) in each monomer, and the 3' G falls into the central pore (Schumacher et al., 2002). Despite the low RMSD of Cα changes (1.4 Å) between apo- and RNA-bound Hfq, significant structural changes are observed in loop 5, which connects β4 and β5 and forms the entrance of the central pore. The structural changes induced by RNA binding cause expansion of the central pore diameter from 12 Å to 15 Å, and they suggested the 3' end of the RNA might thread through the pore (Leung et al., 2011). Later structural studies revealed that RNA does not thread through the central pore. However, they revealed more dynamic interaction through various surfaces.

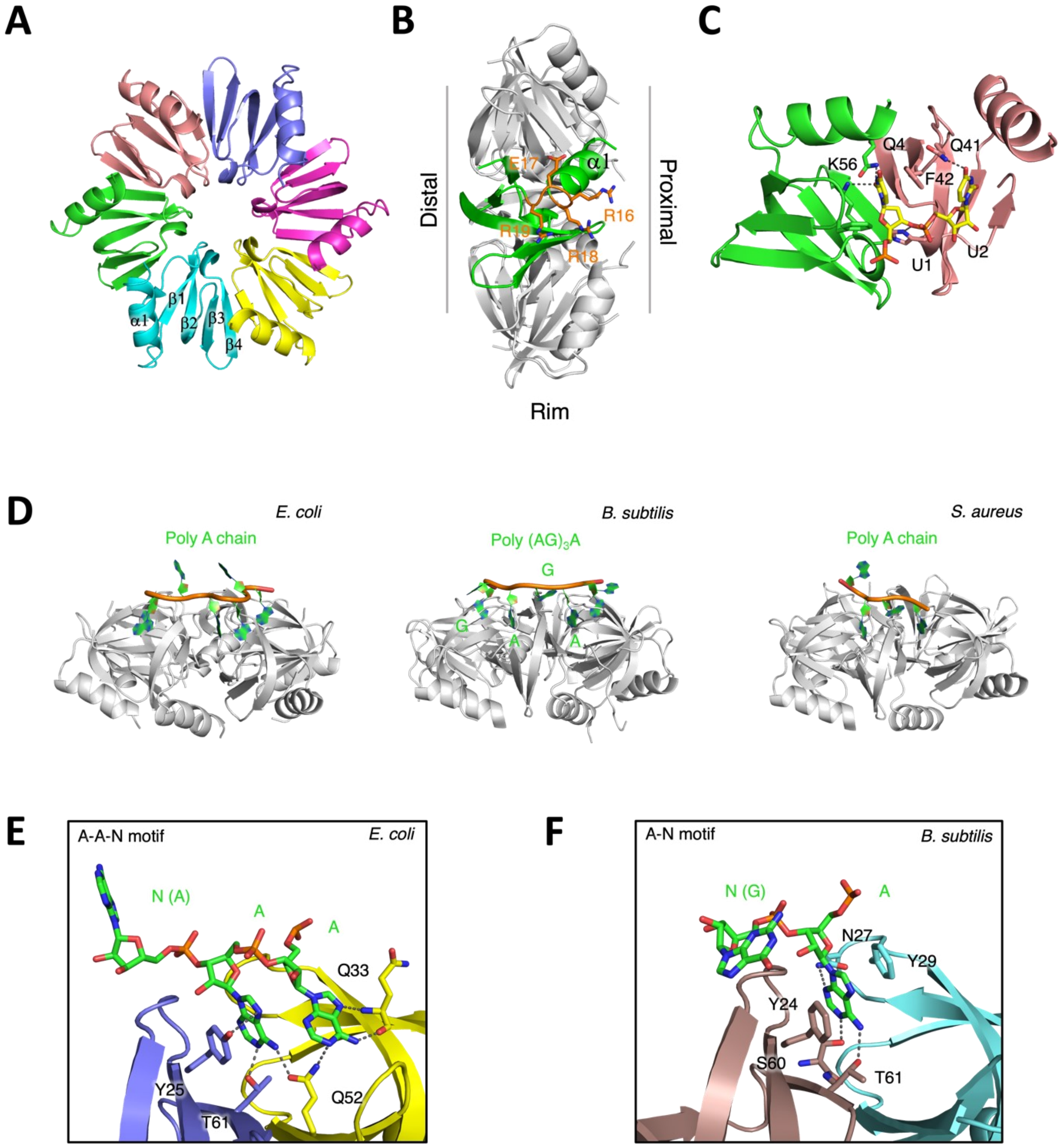


Figure 1-5. Hexameric Hfq has multiple binding sites.

(A) The Hfq monomer, which adopts a $\beta_5\alpha_1$ fold, forms a hexameric ring. (B) Hfq has multiple binding sites, proximal, distal, arginine-rich region ($R_{16}E_{17}R_{18}R_{18}$) on the rim. (C) Molecular interaction of the proximal face with uridines. (D) Various modes of RNA recognition via the distal face of Hfq from the *E. coli, B. subtilis*, and *S. aureus* homologs. (E) Molecular interaction of A-A-N RNA motif with Hfq in *E. coli*. (F) Molecular interaction of A- N RNA motif with Hfq in *B. subtilis*.

1.4.1.1.2. The distal face recognizes poly(A) tracts

The distal face is on the opposite side from the proximal face, and preferentially binds to poly(A) sequences, often found in mRNA 5' UTRs (Link et al., 2009; Soper et al., 2011). Unlike the conserved mode of binding on the proximal face, the distal face showed a variety of modes of RNA recognition (Figure 1-5D). Genomic SELEX identified an enriched motif for Hfq binding, 5'-AAYAAYAA-3' (Y: pyrimidine) (Lorenz et al., 2010). Hfq interacts with A-rich motif through the distal face of protein, and this A-rich motif are different in gram-negative (A-A-N motif, N can be any base) and gram-positive bacteria (A-N motif). Structures of *E. coli* Hfq in complex with poly(A) sequences reveal the molecular mechanism of Hfq A-A-N motif recognition via the distal face (Link et al., 2009; W. Wang et al., 2013). The first adenine forms hydrogen bonds with the main chain NH and CO of Gln33 and the side chain of Gln52 (Figure 1-5E). Tyr25, GlnQ52, and Thr61 recognize the second adenine in the A-A-N motif, and Tyr25 contributes to this interaction through pi-stacking interaction. The N-site can be any base that is flipped out to the opposite side of the protein. On the other hand, the A-N dinucleotide motif is more prevalent for distal face binding in gram-positive bacteria (Someya et al., 2012). Adenine in A-N motif were sandwiched by Tyr24 and Tyr29 in *B. subtilis* (Figure 1-5F). Asn27, Ser60, and Thr61 recognize adenine through hydrogen bonds. The overall mode of binding observed from *B. subtilis* Hfq poly(A) tract showed similar behavior with other gram-positive organisms, such as *S. aureus* Hfq (Horstmann et al., 2012).

1.4.1.1.3. The Hfq rim contributes additional interactions for sRNA binding

An arginine-rich region of the Hfq rim was reported as an additional RNA binding surface (Figure 1-5B). These arginine-rich sequences are located on the $\alpha 1$ helix, and bind RNA AU-rich motifs. A structure of *E. coli* Hfq in complex with RydC sRNA, revealed the interaction of Asn13, Arg16, and Arg17 on $\alpha 1$ mainly through the backbone of U23 and U24 in RydC (Dimastrogiovanni et al., 2014) (Figure 1-5B). Phe39 on $\beta 2$ forms π-stacking interactions with the base of U24. Despite the slightly different arginine-rich sequences, such as RKER in *Pseudomonas aeruginosa*, and RKKR in *A. aeolicus*, Hfq/RNA structures showed common features in that the positively charged residues make similar contacts to RNA and a conserved Phe stacks with uridine (Murina et al., 2013; Stanek et al., 2017). This structural observation suggests that the arginine-rich motif possibly stabilizes the sRNA interaction by providing an additional binding site for the AU-rich sequence often found immediately 5’ to the GC-rich hairpin in the Rho-independent terminator. The study reported that the number of arginines in the arginine-rich sequence strongly correlated with RNA annealing activity (Zheng et al., 2016). Considering the RNA annealing rate of *E. coli* Hfq (RRER in the arginine-rich region) as 1, the annealing rate of *P. aeruginosa* Hfq (RKER), *L. monocytogenes* Hfq (RKEK), *S. aureus* Hfq (KANQ)were 0.1, 0.03, 0.01, respectively (Zheng et al., 2016).

1.4.1.1.4. C-terminal tail

Unlike the structured Sm/Lsm fold, the C-terminus of Hfq is disordered. Even the length of the tail is varied. For example, *E. coli* Hfq has a 38 amino acid tail, while *B. subtilis* Hfq has only a 9 amino acid tail. Santiago-Frangos and his colleagues determined the crystal structure of *Caulobacter crescentus* Hfq with its C-terminus intact (Santiago-Frangos et al., 2019). Interestingly, the DDAD acidic motif, located at the end of the C-terminus, acts like a phosphate backbone that interacts with the neighbouring protomer arginine-rich region on the rim. The *C. crescentus* Hfq C-terminal tail was demonstrated to play an essential role in sRNA selectivity and autoinhibitory activity of RNA annealing.

1.4.1.2. Hfq facilitates RNA chaperone activity through two different mechanisms

Hfq mediates sRNA-mRNA annealing by bringing the interacting RNAs together in close proximity. To mediate annealing, Hfq interacts with both sRNA and mRNA simultaneously using multiple binding sites defined above. Depending on how the RNAs bind to Hfq, mechanisms are categorized into either class I or II (Figure 1-6) (Schu et al., 2015). Class I is the more common sRNA-mRNA annealing mechanism. In this mechanism, the sRNA 3’ poly(U) tail binds to the proximal face and the UA motif 5’ to the hairpin binds the arginine-rich region, located in the rim of Hfq (Figure 1-6A). The mRNA 5' UTR, which contains the AAN motif, binds through the distal face. In the class II mechanism, the sRNA also binds the 3' poly(U) tail through the proximal face, but in this case the 5' AAN motif in the mRNA also interacts through the distal face. Class II mRNAs also have UA motifs within the 5' UTR that interact with the arginine-rich rim (Figure 1-6B).

1.4.2. CsrA, an RNA chaperone that regulates diverse genes through a GGA motif

CsrA was named after the function of the protein, carbon storage regulator (Csr), and regulates glycogen biosynthesis in *E. coli* (Romeo et al., 1993). CsrA regulates gene expression by recognizing GGA sequence motifs, usually within Shine-Dalgarno (SD) sequences, which is a ribosomal binding site in archaeal and bacterial mRNA, as well as coding sequences. Direct CsrA interaction with mRNA often blocks ribosome access to the SD sequence and thereby represses gene expression. High-throughput sequencing and -omics approaches demonstrated that ~800 mRNAs are under the influence of CsrA, accounting for approximately 20% of coding genes in *E. coli* (Edwards et al., 2011; Sowa et al., 2017). One of the genes under the control of CsrA is the RNA chaperone, Hfq. CsrA can downregulate the expression of Hfq by interacting with the SD sequence (UAAGGA) in *hfq* mRNA. An mRNA stability assay showed that CsrA does not affect the half-life of *hfq* mRNA, suggesting CsrA regulates gene expression by competing with 30S ribosomal subunit for SD sequence binding (Baker et al., 2007). Homologs of CsrA, are widely found in *Gammaproteobacteria* and regulate gene regulation at the posttranscriptional level (Pourciau et al., 2020; Quendera et al., 2020).

1.4.2.1. CsrA recognizes GGA motifs in target mRNAs

CsrA is a small protein of approximately 7 kDa in size. CsrA adopts a homodimeric β-sandwhich structure composed of two five-stranded β-sheets followed by a C-terminal α-helix and disordered tail (also known as the RsmA fold; Figure 1-7A) (Gutiérrez et al., 2005). Homologs of *E. coli* CsrA are found in other gram-negative bacteria, such as *Yersinia enterocolitica* (Heeb et al., 2006) and *Pseudomonas fluorescens* (Schubert et al., 2007), and in gram-positive bacteria such as *Geobacillus thermodenitrificans* (Altegoera et al., 2016). The solution structure of CsrA in

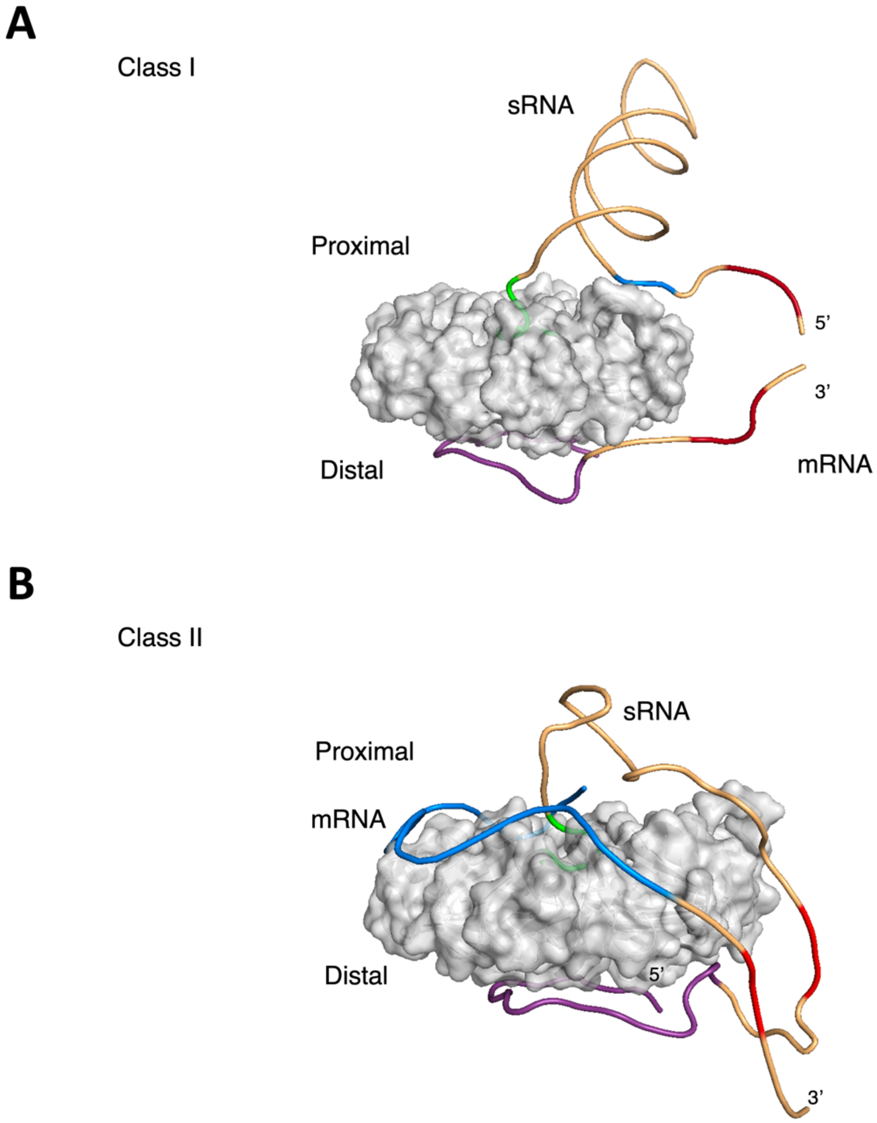


Figure 1-6. Two alternative classes of sRNA-mRNA recognition and annealing by Hfq.

(A) 3D model of the class I mechanism. Poly(U) tract of sRNA (green) binds through the proximal face and additional contacts through the 5’ single-stranded side of the hairpin (blue) through the arginine-rich region of Hfq. A-rich motif of mRNA (purple) binds the distal face of Hfq. Hfq mediates sRNA-mRNA annealing through red strands. (B) 3D model of class II mechanism. U tract and A-rich motif on sRNA interact through both proximal and distal face of Hfq. mRNA wraps Hfq through an arginine-rich motif. Hfq mediates sRNA-mRNA through the red strand. The colour indication is the same as Figure 1-2A.

complex with two target RNAs revealed that one homodimer CsrA can interact with two RNA fragments through GGA motifs (Figure 1-7B) (Schubert et al., 2007). The CsrA structure from *G. thermodenitrificans* showed that CsrA could interact with the FliW protein, preventing CsrA induced translation repression of the flagellar filament protein, flagellin (Altegoera et al., 2016). FliW interacts with CsrA mainly through the extended C-terminus, and the CsrA RNA binding surface also contributed to FliW interaction (Figure 1-7C). Furthermore, the docking model of the trimer complex (CsrA-FliW-*hcnA* mRNA) revealed that FliW probably physically hinders the binding surface, and may also electrostatically repel RNA due to the overall negative charge of FliW (Figure 1-7D) (Altegoera et al., 2016).

1.4.2.2. CsrA regulates gene expression by direct contact with mRNA 5' UTRs

Unlike Hfq, which often regulates gene expression through the action of sRNAs, CsrA regulates mRNAs directly through their GGA sequences. SELEX and CLIP-seq identified that CsrA preferentially binds GGA sequences, and Shine-Dalgarno (SD) sequences often contain GGA sequences (Dubey et al., 2005; Holmqvist et al., 2016; Potts et al., 2017). Direct interaction of the CsrA/RsmA family with mRNA can have a variety of consequences, such as translation repression, translation activation, transcription termination, and RNA stabilization (Pourciau et al., 2020). For example, CsrA binding to the GGA motifs in *glgC* mRNA near its SD sequence results in competition with 30S ribosomal subunit binding and inhibition of translation (Mercante et al., 2009). On the other hand, CsrA interaction with GGA motifs in *ymdA* mRNA destabilizes its 5' UTR hairpin structure, exposing the ribosome binding site to activate translation (Renda et al., 2020). Binding of CsrA with *flhDC* mRNA protected the mRNA from RNase E cleavage that increases its stability (Yakhnin et al., 2013). Certain sRNAs, such as CsrB, CsrC, or RsmZ, which contain multiple GGA motifs, can compete with the CsrA target mRNAs to antagonize CsrA-mediated gene regulation (Valverde et al., 2004; Weilbacher et al., 2003). CsrB and CsrC contains 22 GGA sequences in 350 nucleotides and 9 GGA sequences in 245 nucleotides, respectively, thereby providing a large number of potential binding sites. Most of these GGA sequences are within single-stranded loops and likely act as a CsrA sponge. CsrA has also been shown to regulate expression through other mechanisms such as transcription termination, RNA stability, and translation activation through mRNA remodeling (Pourciau et al., 2020).

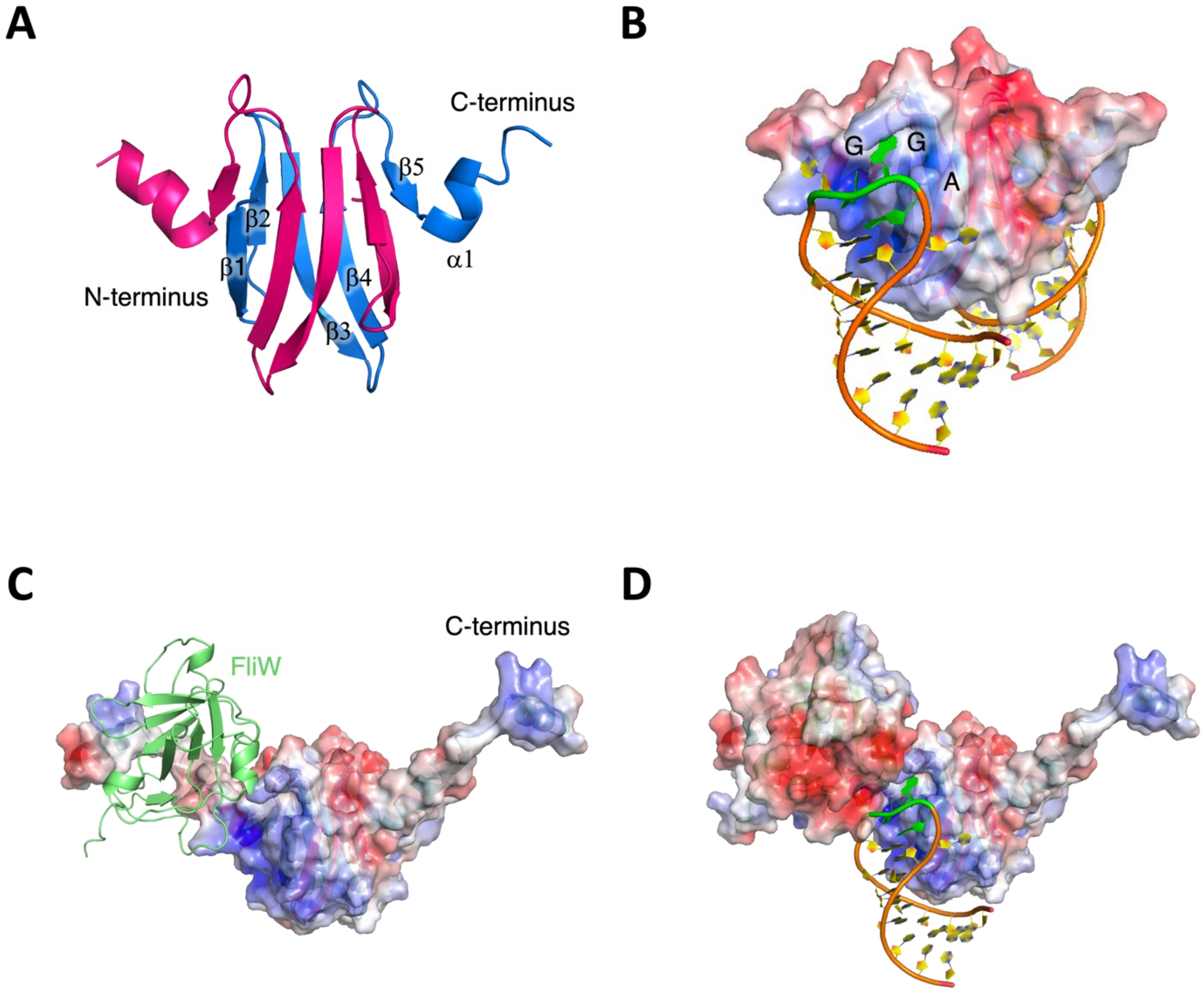


Figure 1-7. Structure of CsrA and interaction with RNA through GGA motif.

(A) Dimeric CsrA structure. (B) CsrA dimer contains two GGA binding sites. (C) CsrA interacts with FliW through an extended C-terminus. (D) Docking model of CsrA/FliW with CsrA/RNA. CsrA and FliW were represented with an electrostatic potential map.

1.5. Discovery of the FinOP system, a repressor of bacterial conjugation

F-plasmid or F factor is a plasmid that can be transferred from donor (F+) to recipient cells (F-). F was first discovered in pioneering work by Lederberg and Tatum nearly 80 years ago and early work on this plasmid was instrumental in revealing the fundamental nature of the genetic material (Tatum & Lederberg, 1946). F factor transfer is mediated by one of the horizontal gene transfer mechanisms, called conjugation, which requires direct contact through a pilus connecting the donor and recipient cells. Many genes are required for pilus formation and most of these are coordinately expressed within a single operon called the *tra* (transfer) operon under the control of the P_γ promoter. The master regulator of the *tra* operon is the TraJ transcription factor that activates the *tra* operon (Achtman, 1973).

Shortly after the discovery of F, a number of related plasmids, called R-factors, were isolated from pathogenic bacteria associated with antibiotic resistance. Unlike F, many of the R-factors showed a much lower transfer frequency. Interestingly, it was shown that an R factor could reduce the transfer of a co-resident F plasmid, suggesting the R factor expressed a transfer repressor (Meynell & Datta, 1967). Finnegan and Willetts discovered two episomal mutants on *traP* and *traO* from these R plasmids and speculated these genes affected TraJ synthesis or function, resulting in conjugation repression (Finnegan & Willetts, 1971). Despite the impact of *finP (traP)* and *finO (traO)* null mutants, the nature of these two genes was unknown. R6-5 plasmid was digested into small fragments using restriction enzymes, EcoRI and HindIII, and cloned into a high copy number plasmid to characterize the genes (Timmis, Cabello, et al., 1978). One of the fragments containing the *finO* gene was successfully translated *in vitro,* and FinO was identified as a 22 kDa protein (Timmis, Andres, et al., 1978). Meanwhile, *finP* was found to encode a small, ~80 nucleotide transcript (Mullineaux & Willetts, 1985). FinP was shown to be antisense to TraJ 5' UTR, and the finding that a mutation in the potential open reading frame (ORF) of *finP* failed to influence FinP function as a repressor argued that this transcript likely acted as an antisense RNA (Koraimann et al., 1991). Later, it turned out that FinO can remodel FinP to promote hybridization with *traJ* mRNA 5' UTR to downregulate the expression of the master conjugal regulator, TraJ protein (Arthur et al., 2003).

1.5.1. Characterization of FinP, FinO, and its interaction

FinP is encoded from the complementary strand of the *traJ* mRNA 5' UTR region and the folds of two transcripts were predicted to be similar. Partial RNase digestion was used to probe the secondary structure of *in vitro* transcribed 79 nucleotide FinP and *traJ* mRNA 5' UTR (van Biesen et al., 1993). Enzymatic mapping demonstrated that FinP contains two stem-loops, SLI and

SLII, as predicted in previous studies (Finlay, B. B. et al., 1986; Koraimann et al., 1991), while mapping of *traJ* mRNA 5' UTR revealed three stem-loops, SLIII, SLIIc, and SLIc. SLIc includes the ribosome binding site and AUG start codon. SLIIc and SLIc are the complementary sequences to SLII and SLI in FinP. SLIII is upstream of SLIIc and SLIc.

The structural understanding of FinP and *traJ* mRNA enabled the investigation of interactions with its binding partner, the FinO protein. Direct binding of FinO to FinP and *traJ* mRNA was demonstrated and deletion analysis showed that this interaction is via SLII, not SLI (Biesen & Frost, 1994). Jerome and colleagues focused on the structural differences in SL1 and SLII and tested the importance of different structural elements in SLII for FinO binding using gel-shift analysis. This work demonstrated that FinO recognizes the hairpin as well as the 5' and 3' single-stranded tails in a sequence independent manner (Jerome & Frost, 1999). Limited proteolysis of FinO protein both alone and in presence of FinP SLII revealed two RNA binding regions of FinO, the N-terminus (26-61) and ProQ/FinO domain (62-186) (Ghetu et al., 1999). The crystal structure of $\text{FinO}_{26\text{-}186}$ revealed a shape resembling a fist with an extended helical "index finger" (Figure 1-8A). The structure overall is comprised of a highly alpha-helical fold, agreeing with circular dichroism measurements (Ghetu et al., 1999). Positively charged surfaces on the palm side of the fist (concave surface, residues 62-186) and extended index finger (26-61) supported the idea that SLII of FinP binds to FinO through two different regions (Figure 1-8B). This hypothesis was tested using protein-RNA UV-induced crosslinking and FRET (Ghetu et al., 2002). Analysis of 12 single substitution cysteine mutants indicated the most efficient crosslinking was on the positively charged, concave patch. For FRET experiments, a two stranded version of SLII was created which lacked the loop and the two separate strands were 5' labeled with fluorescein which served as the donor fluorophore. The cysteine substitutional mutants prepared for crosslinking were individually labeled with an acceptor fluorophore, Texas red. The results suggested that FinO contacts SLII of FinP through the single stranded regions and the bottom of stem. RNase footprinting showed FinO protected the 3' tail and base of the stem, however the 5' tail was not protected. Interestingly, protection of the stem was different in that the 3' side of stem was more protected than the 5' side of the stem. Crosslinking, gel-FRET, and RNase footprinting experiments provided biochemical restraints, which, together with SAXS data, were used to model the FinO:SLII structure (Arthur et al., 2011). The SLII base and 3' tail were modeled near the concave surface on ProQ/FinO domain and the stem of SLII was positioned along the extended $\alpha 1$ of FinO (Figure 1-8B).

The FinO bound to FinP complex was believed to play important role in hybridizing FinP and *traJ* mRNA to repress bacterial conjugation (Mark Glover et al., 2015). The effects of point mutations in FinP and *traJ* mRNA on conjugation rates suggested that FinP-*traJ* association initiates through interactions between the complementary loops in a "kissing complex" (Gubbins et al., 2003; Koraimann et al., 1991). The interaction of FinO with FinP was reported to enhance the pairing of FinP with *traJ* mRNA *in vitro* (Ghetu et al., 2000). Duplex pairing rates were heavily influenced by the apical 25 residues at the N-terminus of FinO in that hybridization efficiency of the two RNAs with $FinO_{26-186}$ was decreased 10-fold compared to FinO full-length (Ghetu et al., 1999, 2000). The functional role of the FinO N-terminus was clarified through RNA stability and mating efficiency assays (Arthur et al., 2003). $FinO_{26-186}$, $FinO_{45-186}$, and FinO full-length showed a similar level of RNA protection in the RNA stability assay, however, the conjugation repression efficiency of the N-terminal truncation mutants, $FinO_{26-186}$, $FinO_{45-186}$, compared to FinO full-length decreased 17-fold and 20-fold, respectively. These results demonstrated that the ProQ/FinO domain of C-terminal FinO is essential for RNA binding, however the N-terminal region is essential for chaperone activity to facilitate RNA pairing to repress bacterial conjugation. However, the mechanism by which the N-terminal region facilitates RNA association is unknown.

1.5.2. FinO is an RNA chaperone that rearranges its target RNAs and promotes duplex formation

The chaperone mechanism of FinO was tested through a strand exchange assay and an RNA annealing assay. RNA chaperones reshape RNA structure that can be thought of as two complementary processes: destabilization of RNA secondary structures and facilitation of new created by essentially deleting the hairpin loop (Figure 1-9A). A duplex consisting of a 5' radiolabelled strand SII(A) and non-labeled strand SII(B) was incubated with excess of non-labeled SII(A) in the presence of various FinO mutants to assess the strand exchange ability of FinO as a function of time as visualized by the conversion of the labeled duplex to a single strand by gel electrophoresis (Figure 1-9B) (Arthur et al., 2003). No strand exchange activity was

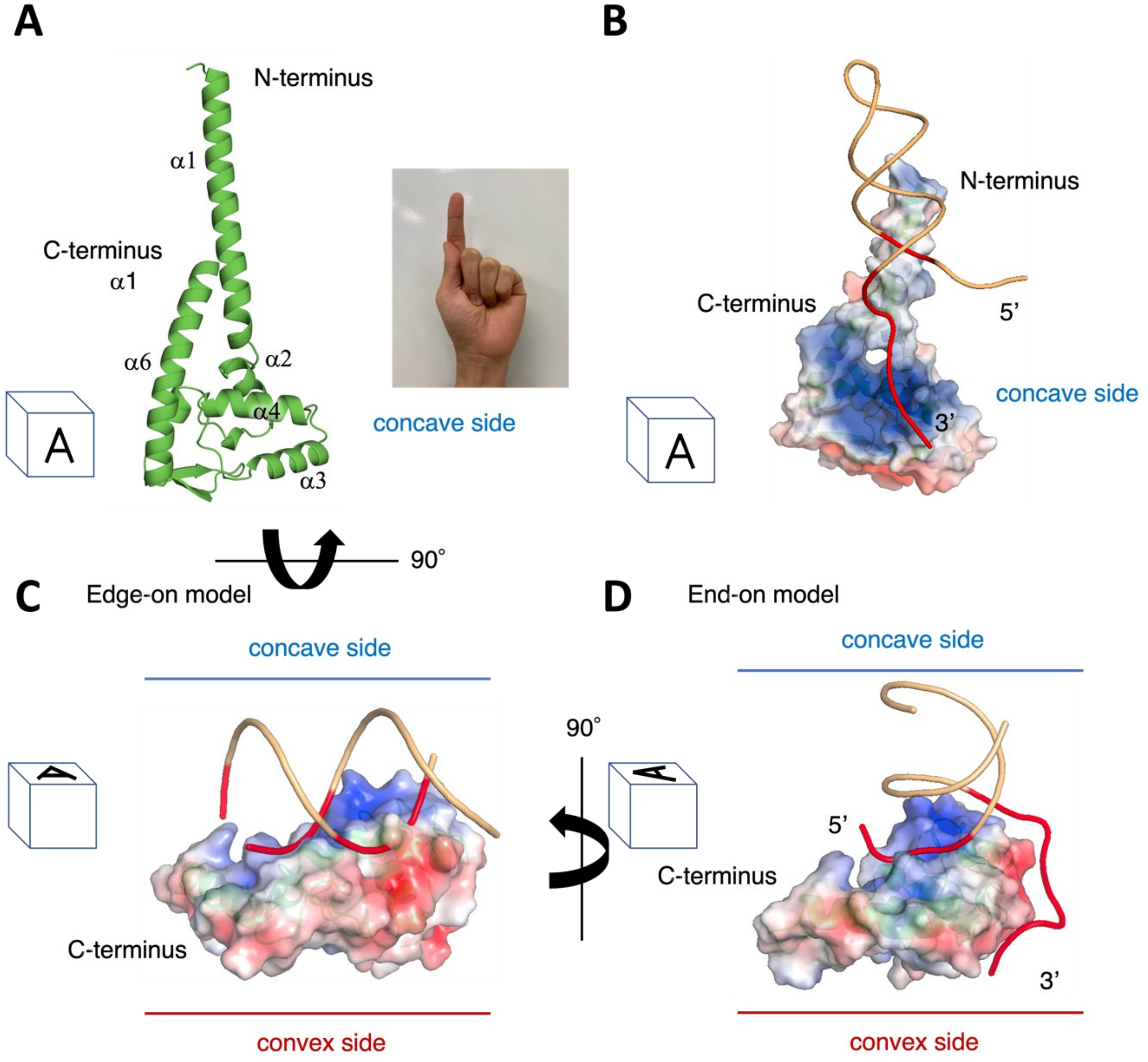


Figure 1-8. ProQ/FinO domain structure and predicted interaction with target RNA.

(A) The crystal structure of $\text{FinO}_{26\text{-}186}$ resembles a left hand. A on cube indicating which side you are looking at. (B) FinO/FinP model derived from integrative structural modeling. The red part of RNA is where FinP is protected from RNase by FinO (Arthur et al., 2011). (C) The edge-on model with the NMR ProQ structure with A-form RNA helix the region contacting the protein in red. Current view is bottom view from (A) (Pandey et al., 2020). (D) The end-on model with the NMR ProQ structure and an A-form RNA helix with 3' and 5' tails. RNA regions in contact with protein are in red (Pandey et al., 2020).

observed in the absence of FinO over a 2 hour time course, indicating that SII(A) and SII(B) form base pairing interactions. The ability to destabilize internal base-pairing was assessed using the strand exchange assay. For this experiment, a two stranded version of FinP SLII was a stable duplex. FinO, however, was able to facilitate complete strand exchange within 2 hours, suggesting a dramatic enhancement of exchange. Intriguingly, the N-terminal disordered region and α1 were both essential to this exchange activity, even though deletion of these regions did not impair RNA binding activity. Further mutagenesis of this N-terminal region revealed the N-terminus of α1 as being particularly important for strand exchange.

The RNA annealing assay used radiolabeled SLII and an excess of the complementary non-labeled SLIIcx. Similarly to the duplexing assay, interactions between the RNAs were monitored by gel electrophoresis. In the absence of FinO, the two hairpins were not observed to associate over a 2 hour incubation, however, FinO facilitated duplexing of these RNAs within just a few minutes. Again this activity was dependent on the N-terminus as duplexing was retarded with deletion or point mutations in the FinO N-terminus. A cartoon of the proposed function of FinO to facilitate FinP-*traJ* interactions is presented in Figure 1-10 (Mark Glover et al., 2015).

1.5.3. The FinO family is distributed across proteobacteria

Phylogenetic studies mapped a wide distribution of ProQ/FinO domain containing proteins across γ-proteobacteria (Attaiech et al., 2016; Katsuya-gaviria et al., 2022; Olejniczak & Storz, 2017; Quendera et al., 2020). The first two FinO family proteins to be studied were ProQ in *E. coli* (Chaulk et al., 2011; Kunte et al., 1999; Smith et al., 2004), and NMB1681 in *N. meningitidis* (Chaulk et al., 2010), and this early work showed that both these proteins could, like FinO, bind transcriptional terminator structures and also showed some strand exchange and RNA duplexing activity. Intriguingly, 674 ProQ/FinO domains were found in 553 proteomes, indicating some strains contain more than one ProQ/FinO gene in their genome (Attaiech et al., 2016). For example, *L. pneumophila* contains two ProQ/FinO homologs, RocC and Lpp1663. RocC consists of 230 amino acids having a ProQ/FinO domain in the N-terminus, and a disordered C-terminus. On the other hand, the second protein, Lpp1663, lacks the extended disordered C-terminus and is essentially only a FinO domain (Attaiech et al., 2016; Immer et al., 2020). The fact that FinO proteins are often plasmid encoded raises the possibility of multiple FinO proteins in the same cell. For example, *S. enterica* strain SL1344 harbors three species of FinO family members:

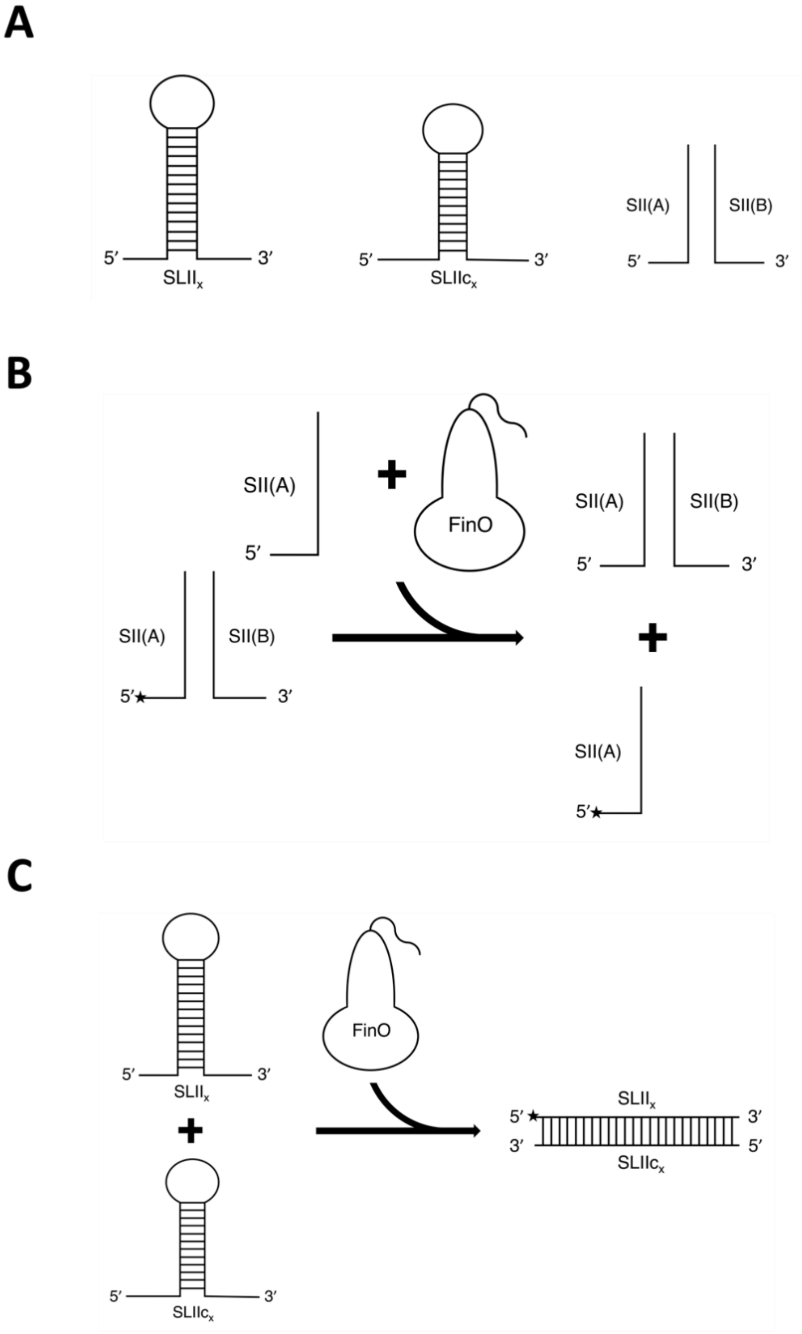


Figure 1-9. Schematic diagram of strand exchange and annealing assays (Image was taken from (H. J. Kim et al., 2020)**).**

(A) Set of RNA constructs for RNA annealing and strand exchange assays. SLII_x and SLII_{cx} are complementary RNAs. SII (A) and SLII (B) pair to form a duplex structure similar to SLIIxc but lacking the loop. (B) Schematic diagram of RNA annealing assay. Star indicates radiolabeled. (C) Schematic diagram of strand exchange assay.

chromosomally-encoded ProQ and plasmid-encoded FinO and another FinO protein, FopA, encoded on a second plasmid. RIP-seq showed distinct targets for FinO, such as FinP and RepX, and ProQ, suggesting they regulate different sets of genes (El Mouali et al., 2021). In the following sections, I will discuss three of the more studied proteins: *E. coli* ProQ, *N. meningitidis* NMB1681, and *L. pneumophila* RocC.

1.5.3.1. ProQ, a global RNA binding protein via interaction with structured sRNAs and mRNAs

E. coli ProQ was known as a regulator of ProP, which is an osmosensor and osmoregulatory transporter. Modeling suggested that ProQ contains an N-terminal FinO domain and a C-terminal Tudor/Hfq-like domain (Chaulk et al., 2011) and the structures of these domains were later confirmed by NMR (Gonzalez et al., 2017). Early work demonstrated that the ProQ FinO domain could bind hairpin-3' tail transcription terminator structures and the C-terminal Tudor domain also had some RNA binding activity. Like FinO, ProQ could also catalyze strand exchange and RNA duplexing, suggesting it might act as an RNA chaperone (Chaulk et al., 2011). However, more a definitive understanding of ProQ function required the elucidation of its true *in vivo* RNA targets.

The advent of the high-throughput sequencing of bacterial genomes changed the understanding of FinO family binding partners. Grad-seq revealed diverse RNA binding partners of ProQ by partitioning cellular transcripts based on glycerol gradient which were subsequently sequenced. The analysis of sequenced RNAs revealed that ProQ interacts with hundreds of RNAs including many non-coding sRNAs and mRNAs in *Salmonella enterica* (Smirnov et al., 2016). ProQ deletion was found to affect the expression of ~800 transcripts, ~ 16% of genome, suggesting the importance of ProQ-dependent gene regulation in *S. enterica*. UV-CLIP coupled with RNA sequencing suggested that ProQ pervasively binds to highly structured regions of RNA, such as mRNA and sRNA 3' terminators (Holmqvist et al., 2018). The *cspE* mRNA 3' end is one of the ProQ targets containing a transcriptional terminator. The binding of ProQ with the *cspE* 3' end increased the stability of this RNA, particularly against the 3' – 5' exoribonuclease RNAse II, similar to how FinO protects FinP from RNase E degradation. The transcriptional terminator

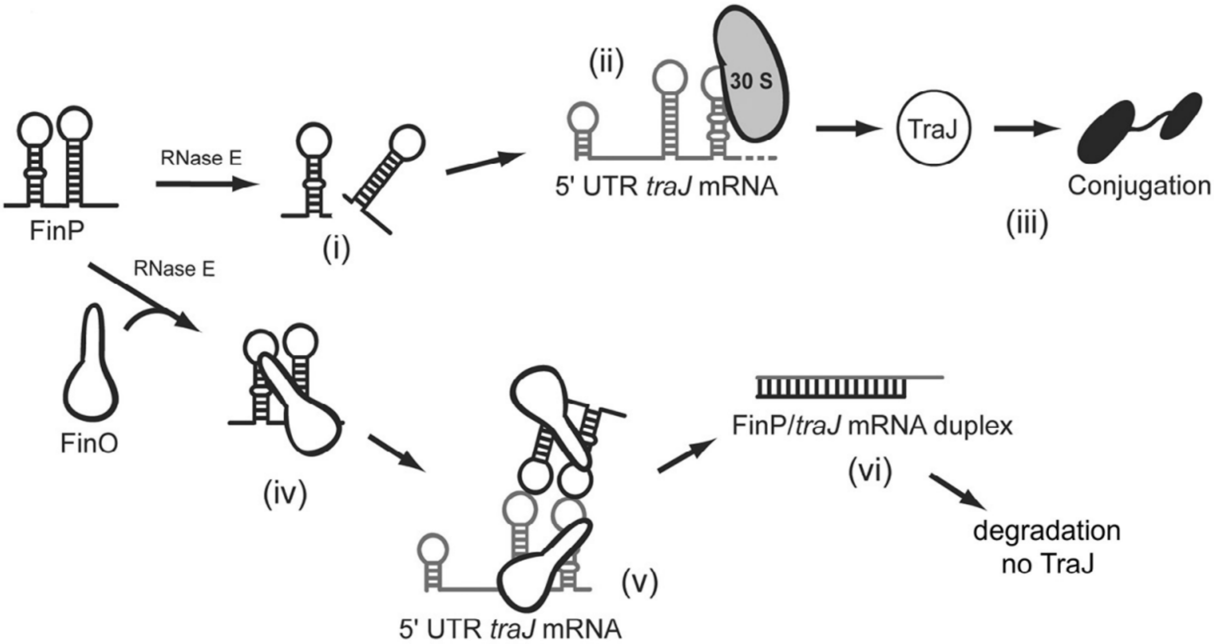


Figure 1-10. Predicted mechanism for how the FinOP system regulates bacterial conjugation (Image was taken from (Mark Glover et al., 2015)**).**

(i) FinP is digested by RNase E in absence of FinO. (ii) 30S ribosomal subunit binds and initiates translation of TraJ mRNA. (iii) TraJ activates *tra* operon transcription required for bacterial conjugation. (iv) FinO binding protects FinP from RNase degradation. (v) FinO facilitates FinP and *traJ* mRNA kissing interactions (vi) FinP/*traJ* mRNA duplex formation inhibits TraJ translation.

was dissected into loop, stem and tail and these were investigated to identify the minimal determinants recognized by ProQ using EMSA. This revealed that the minimum requirement to bind to ProQ is a 2 base-pair stem with a 4 nucleotide 3' tail, with no apparent sequence specificity (Stein et al., 2020).

Discovery of biologically relevant RNA targets also enabled the mapping of important residues on the surface on ProQ using random mutagenesis. Pandey and colleagues screened the effect of mutations on ProQ/FinO binding to the RNA ligands, *cspE* mRNA and SibB (Stein et al., 2020). They found that a collection of positively charged and other residues on the concave surface impaired RNA binding, similar to the earlier results from protein-RNA crosslinking in the FinO system. However, they suggested two quite different possible modes of RNA interaction for *E. coli* ProQ than had been previously modeled for FinO (Pandey et al., 2020). They focused on the fact that critical residues for binding to RNA were periodically distributed 15-20Å, which is close to the width of an A-form helix. They suggested two possible models. The first they called the edge-on model in which an A-form helix is rotated ~120° compared to the FinO/FinP model, to give extensive potential electrostatic interactions between the negatively charged phosphate backbone and the positively charged ProQ/FinO surface (Figure 1-8C). The second mode they called the "end on" model, in which the base of the stem-loop is docked against the positive surface, while the 3' tail wraps around the convex side of the protein (Figure 1-8D). The end-on model can explain protection of the 3' tail, which the edge-on model could not. Thus, in spite of the large base of biochemical and biophysical data, a clear consensus structural picture for how these proteins bind RNA was lacking. Saturation mutagenesis confirmed the residues on the ProQ/FinO domain that abolish ProQ-dependent gene regulation. The large overlap in the mutation library in these two mutagenesis studies showed a strong relationship between RNA binding and regulatory activity (Rizvanovic et al., 2021).

1.5.3.2. NMB1681, a minimal ProQ/FinO protein with many RNA targets

NMB1681 is a minimal FinO family member that lacks both N- and C-terminal extensions. The crystal structure of NMB1681 showed high degree of structural similarity with FinO family crystal structure (RMSD: 1.0 Å). However, NMB1681 showed a somewhat different charge distribution compared to FinO and ProQ. ProQ and FinO have a major positively charged surface on the concave side, however, NMB1681 has strong positive charge surfaces on the convex side (Chaulk et al., 2010; Olejniczak & Storz, 2017). In spite of these differences, NMB1681 was shown to bind model transcriptional terminators and also exhibited RNA strand exchange and duplexing activities, albeit with a lower strand exchange activity compared to FinO (Chaulk et al., 2010). NMB1681 could partially repress F conjugation in the absence of FinO (Chaulk et al., 2010). Insights into the true physiological targets of NMB1681 were revealed through UV-CLIP combined with RNA sequencing, showing that NMB1681 can bind to 16 sRNAs and 166 mRNAs, mainly through their 3' UTRs (Bauriedl et al., 2020). UV crosslinking-induced mutations also showed that NMB1681 binding sites are mainly within structured regions of the RNA targets. The binding of NMB1681 to target mRNAs resulted in strong inhibition of digestion by PNPase, which is a 3' exoribonuclease. They furthermore showed that NMB1681 is essential for oxidative stress tolerance and DNA damage repair in *N. meningitidis*.

1.5.3.3. RocC, a specific RNA binding protein that regulates competence development in *L. pneumophila*

In certain growth conditions, *Legionella pneumophila* can be competent to uptake exogenous DNA from the environment. Competence development requires expression of a set of genes (the competence regulon) that establish the DNA transport pore and other accessory proteins. Deletion studies demonstrated two key components in this regulation, RocC, a ProQ/FinO domain containing protein, and RocR, a small non-coding RNA. These together act to repress the competence regulon, containing *comEA, comEC, comM,* and *comF* (Attaiech et al., 2016). RocR consists of three stem-loops and the first two, SL1 and SL2, were predicted to imperfectly pair the with 5' UTRs of the competence regulon RNAs through a target seed sequence called the RocR box. EMSA demonstrated the specific recognition of the transcriptional terminator SL3 by ProQ/FinO domain of RocC and that mutation of the ProQ/FinO domain abrogates RocC-mediated repression.

1.6. Organization of the thesis

The presented work attempts to elucidate the molecular mechanism of how ProQ/FinO domains recognizes the intrinsic terminator RNA structures. To achieve this goal, we chose the highly specific ProQ/FinO domain of RocC and its target RNA, RocR, from *L. pneumophila*. A previous preliminary study showed that the ProQ/FinO domain of RocC binds to SL3 of RocR, the transcriptional terminator, with a similar binding affinity as intact RocC and RocR (Attaiech et al., 2016). Chapter 3 describes the critical regions of $\mathrm{RocR_{SL3}}$ and RocC that are required for recognition. We truncated $\mathrm{RocC_{1\text{-}126}}$ from the N-terminus and revealed that $\mathrm{RocC_{14\text{-}126}}$ is the minimal RocC that tightly binds to SL3 via EMSA. Various truncation and substitution mutants of $\mathrm{RocR_{SL3}}$ were tested to define the determinants in $\mathrm{RocR_{SL3}}$ that are essential for RocC recognition through EMSA. Chapter 4 describes our structure of a RocC/RocR complex, as well as the methods we developed to express and purify RocC, RocR and their complexes, and how we eventually crystallized and determined the structure of a RocC/RocR complex. Chapter 5 describes structure-guided mutagenesis studies with binding and *in vivo* assays that explore the importance of the structure for RocC-RocR binding *in vitro*, and for RocC function *in vivo*. This chapter also provides evidence that RocC and other ProQ/FinO domains bind transcription terminators in a manner that is sensitive to the length of the 3’ ssRNA tail. Chapter 6 summarizes the overall work and discussed future directions. Appendix A describes preliminary experiments to understand the chaperone activity of RocC. Appendix B describes the dynamic interaction of iASPP-PP1c using SEC-SAXS.

Chapter 2.

Materials and Methods

2.1. Cloning RocC truncation mutant and site directed-mutagenesis for *in vitro* assays

RocC mutants used in this study were amplified from pET-47b(+) RocC$_{1\text{-}230}$ (Attaiech et al., 2016) using appropriate primers purchased from IDT (Table 2-1 and 2-2) and the Phusion high-fidelity DNA polymerase (NEB). The resulting *rocC* gene fragments were digested and ligated between the BamH1 and Not1 sites in pGEX-6P-1 such as to produce a GST-(HRV-3C)-RocC protein. Ligated DNAs were transformed into E. coli DH5α (Invitrogen). Transformants were selected on LB plates with ampicillin. Obtained plasmids were verified using Sanger sequencing.

2.2. Protein expression and purification

For each RocC, a single colony of BL21-Gold (DE3) bearing one of the pGEX-6P-1_RocC plasmid was inoculated into 25 mL LB with 100 mg/mL ampicillin and 35 mg/mL kanamycin and incubated at 37°C for 18 hours with shaking. This preculture was then inoculated into 1L LB with 100 mg/mL ampicillin and 35 mg/mL kanamycin and grown to O.D. 0.6-0.8 at 37°C. GST-PP-RocC expression was induced with 0.3 mM IPTG and incubation was continued at 18°C for 20 hours. The cell pellet was collected by centrifugation, flash frozen and stored at -80°C. The cell pellet was resuspended in lysis buffer (50mM HEPES pH 7.3, 500 mM NaCl, 5% glycerol, 1 mM DTT) and lysed using an Emulsiflex-C3 high pressure homogenizer (Avestin). The lysate was incubated with 10 mL glutathione beads for 1 hour at 4°C. Incubated beads were washed with 150 mL of lysis buffer and the GST tag was removed by digestion with 3C protease for 18 hours at 4°C. The digested protein was collected and purified by Superdex 75 16/60 gel filtration chromatography in different buffers depending on the purpose (ITC/EMSA/FP binding assay buffer: 25 mM pH 7.3 HEPES, 150 mM NaCl, 10% glycerol; apo-protein crystallization buffer: 25 mM Tris pH 8.0, NaCl 30mM, 1 mM DTT; complex crystallization buffer: 10 mM HEPES-KOH pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 1 mM TCEP). FinO$_{45\text{-}186}$ (Arthur et al., 2003) and ProQ$_{1\text{-}130}$ (Chaulk et al., 2011) were purified in the similar way as RocC.

2.3. RNA synthesis, expression, and purification

RNAs were prepared in three ways for different analyses. RNAs for EMSA were produced using *in vitro* transcription (Table 2-3), purified using denaturing PAGE, and radiolabeled as previously described (H. J. Kim et al., 2020). Given templates in Table 2-3 included +1 or +2 guanosine in 5' side of template strand to increase the yield of in vitro transcribed RNAs (Imburgio et al., 2000). RNAs for NMR spectroscopic applications were produced by solid phase synthesis and ^{15}N labeled nucleotides for assignment purposes were

Table 2-1. Oligonucleotides used to create pGEX-6P-1 derivatives with truncation rocC mutants.

Primer	**Sequence (5'-3')**
RocCdel1_Fwd_BamH1	CTAGTCGGATCCATGAGAAAGCAGGCGCT
RocCdel14_Fwd_BamH1	TCAGGATCCAATAAAGCACAAAAAAATCAATCCAAGC
RocCdel24_Fwd_BamH1	CTGAATGGATCCGCGCGATCTGAC
RocC126_Rev_Not1	GACTAGGCGGCCGCTTACTTTTCCACGCGTTTTTTAATTTTC
RocC137_Rev_Not1	GACTAGGCGGCCGCTTATGCATTCACTTGTTTGCGAG

Table 2-2. Oligonucleotides used to create pGEX-6P-1 derivatives with punctual RocC mutants.

Mutation	Forward primer	Reverse primer
	sequence (5'-3')	sequence (5'-3')
K15D	TCAGGATCCAATGATGCACAAAAAAATCAATCCAAGC	-
Q17A	TAAAGCAGCAAAAAATCAATCCAAGCGCG	TTGGATCCCAGGGGCCCC
K18D	TCAGGATCCAATAAAGCACAAGATAATCAATCCAAGC	-
S21A	TCAGGATCCAATAAAGCACAAAAAAATCAAGCCAAGC	-
I51A	GCCATTAAAGGCTGGTATTATGTCTGATATATTG	CGAATACGCAAAGAATTATC
S70A	AGTTGGAGTTGCTAAAAGCAAATTAAG	TGCTCTGCTTTTTCTGCATG
S72A	AGTTTCTAAAGCCAAATTAAGGGAAGCTGTTG	CCAACTTGCTCTGCTTTTTC
S70A, S72A	AAGCCAAATTAAGGGAAGCTGTTG	TAGCAACTCCAACTTGCTCTG
K71D	TGGAGTTTCTGATAGCAAATTAAGGG	ACTTGCTCTGCTTTTTCTG
K73D	TTCTAAAAGCGATTTAAGGGAAGCTGTTG	ACTCCAACTTGCTCTGCT
R75D	AAGCAAATTAGATGAAGCTGTTGTTCTTTTTAC	TTAGAAACTCCAACTTGC
T82A	TGTTCTTTTTGCCCGTCGTCTTG	ACAGCTTCCCTTAATTTGC
R83D	TCTTTTTACCGATCGTCTTGATTATCTGGC	ACAACAGCTTCCCTTAATTTG
Y87F	CGTCTTGATTTTCTGGCTTGC	ACGGGTAAAAAGAACAAC
R97M	TCGTGAAGTCATGATCGATTTGCATGGAAATCCAGTAGC	GCTTTCAGGCAAGCCAGA
N115A	AGAAGCGGAGGCTGCTTCCATGAAAATTAAAAAAC	TCCTCAGTAACCTCCGCT
K119D	-	GACGCGGCCGCTTACTTTTCCACGCGTTTTTTAATATCCAT

Table 2-3. DNA templates used for in vitro transcription.

DNA template	Sequence (5'-3')
T7 RNAP promoter Rev	TAATACGACTCACTATAGG
RocR$_{\text{SL3}}$ Fwd	AGAAAGGGCCAATCAGTGTGTCGCCAATTGACCCACCCTATAGTGAGTCGTATTA
RocR$_{\text{SL3}}$ Rev	TAATACGACTCACTATAGGGTGGGTCAATTGGCGACACACTGATTGGCCCTTTCT
RocR$_{\text{3nt}}$ Fwd	AAAGGGCCAATCAGTGTGTCGCCAATTGACCCACCCTATAGTGAGTCGTATTA
RocR$_{\text{3nt}}$ Rev	TAATACGACTCACTATAGGGTGGGTCAATTGGCGACACACTGATTGGCCCTTT
RocR$_{\text{4nt}}$ Fwd	GAAAGGGCCAATCAGTGTGTCGCCAATTGACCCACCTATAGTGAGTCGTATTA
RocR$_{\text{6nt}}$ Fwd	AAGAAAGGGCCAATCAGTGTGTCGCCAATTGACCCACCTATAGTGAGTCGTATTA
RocR$_{\text{7nt}}$ Fwd	AAAGAAAGGGCCAATCAGTGTGTCGCCAATTGACCCACCTATAGTGAGTCGTATTA
RocR$_{\text{stem}}$ Fwd	AGAAAAAAACAATCAGTGTGTCGCCAATTGTTTTACCTATAGTGAGTCGTATTA
RocR$_{\text{AU}}$ Fwd	ATAAAGGGCCAATCAGTGTGTCGCCAATTGACCCACCCTATAGTGAGTCGTATTA
RocR$_{\text{AU}}$ Rev	TAATACGACTCACTATAGGGTGGGTCAATTGGCGACACACTGATTGGCCCTTTAT
RocR$_{\text{CA}}$ Fwd	TGAAAGGGCCAATCAGTGTGTCGCCAATTGACCCACCTATAGTGAGTCGTATTA
RocR$_{\text{9bp}}$ Fwd	AGAAAGGGCCAATCTGTGTCGAATTGACCCACCTATAGTGAGTCGTATTA
RocR$_{\text{7bp}}$ Fwd	AGAAAGGGCCAATGTGTCGTTGACCCACCTATAGTGAGTCGTATTA
RocR$_{\text{7bp}}$ Rev	TAATACGACTCACTATAGGTGGGTCAACGACACATTGGCCCTTTCT
RocR$_{\text{5bp}}$ Fwd	AGAAAGGGCCTGTGTCGGACCCACCTATAGTGAGTCGTATTA
RocR$_{\text{5bp}}$ Rev	TAATACGACTCACTATAGGTGGGTCCGACACAGGCCCTTTCT
RocR$_{\text{3bp}}$ Fwd	AGAAAGGGTGTGTCGCCCACCTATAGTGAGTCGTATTA
RocR$_{\text{3bp}}$ Rev	TAATACGACTCACTATAGGTGGGCGACACACCCTTTCT
RocR$_{\text{11bp-tet}}$ Fwd	AGAAAGGGCCAATCACCGAAGCAATTGACCCACCCTATAGTGAGTCGTATTA
RocR$_{\text{11bp-tet}}$ Rev	TAATACGACTCACTATAGGGTGGGTCAATTGCTTCGGTGATTGGCCCTTTCT
RocR$_{\text{9bp-tet}}$ Fwd	AGAAAGGGCCAATCCGAAGATTGACCCACCCTATAGTGAGTCGTATTA
RocR$_{\text{9bp-tet}}$ Rev	TAATACGACTCACTATAGGGTGGGTCAATCTTCGGATTGGCCCTTTCT

* Constructs does not have reverse DNA used T7 RNAP promoter Rev as a Reverse template.

incorporated as described earlier (Plangger et al., 2019). FAM-labeled RNAs for FP and RNAs for ITC were either purchased from IDT or synthesized. RNAs for crystallization were prepared using anion exchange followed by gel filtration, as previously described (H. J. Kim et al., 2020).

2.4. Limited proteolysis

Limited proteolysis with different RocC:protease ratio

10 μL of 5 mg/mL RocC$_{1\text{-}230}$ in the superdex buffer (25 mM Tris pH 8.0, 150 mM NaCl, 10% Glycerol, 1 mM DTT) was mixed with 0.01 mg/mL trypsin and chymotrypsin (Proti-Ace kit, Hampton research) to make RocC:protease ratio from 100:1 to 5,000:1 (w/w). Leftover volume was filled up with Proti-Ace dilution buffer (10 mM HEPES pH 7.5, 500 mM NaCl) to 20 μL. The

reaction was incubated for 45 minutes at 37°C. Taken samples at each time point were mixed with x5 SDS sample buffer to load on the 15% SDS-PAGE gel.

Time course of trypsin digestion at a 2,500:1 ratio

The trypsin digestion reaction was mixed with 10 μL of 4 mg/mL $\mathrm{RocC}_{1\text{-}230}$ in the superdex buffer was incubated with and 1.6 μL of 0.01 mg/mL trypsin (Proti-Ace kit, Hampton research), 8.4 μL of Proti-Ace dilution buffer. The mixture of the 2,500:1 ($\mathrm{RocC}_{1\text{-}230}$: trypsin) was incubated at 37°C. Taken samples in each time point was mixed with x5 SDS sample buffer to load on the 15% SDS-PAGE gel.

2.5. Electrophoretic mobility shift assay (EMSA)

In vitro transcribed RNAs were 5' radiolabeled with ATP, [γ-^{32}P] (PerkinElmer) using T4 polynucleotide kinase (Invitrogen) and purified using denaturing PAGE as described(H. J. Kim et al., 2020). Labeled RNAs were incubated 30 minutes on ice with proteins from 0 to 32 μM concentration in a final volume of 5 μL EMSA reaction buffer (25 mM HEPES pH 7.3, 150 mM NaCl, 4 mM $\mathrm{MgCl_2}$, 10% glycerol, 1 mM DTT, 0.5 mg/mL yeast tRNA (ThermoFisher), and 12 U RNaseOUT (ThermoFisher)). The reactions were mixed with 5x native gel loading dye (10 mM Tris pH 8.0, 50% glycerol, 0.001% bromophenol blue, 0.001% Xylene cyanol FF) and separated on 10% native gels run in 1x Tris-Glycine buffer at 4°C. Gels were dried and imaged with phosphor imaging screens (Molecular Dynamics). Band intensities were quantified using Molecular Dynamics ImageQuaNT TL software (GE healthcare). Prism5 (GraphPad) was used for graph fitting and equilibrium dissociation constants (K_D) calculation with one site – specific binding with Hill slope:

$$Y = \frac{B_{max} \times X^h}{({K_D}^h + X^h)}$$

Where X is the concentration, B_{max} is the maximum specific binding and h is the Hill slope.
All EMSA results were repeated three times.

2.6. SEC-MALS

All samples were prepared in crystallization buffer (10 mM HEPES-KOH pH 7.3, 100 mM KCl, 5 mM $\mathrm{MgCl_2}$, 1 mM DTT) with HPLC grade water (Thermofisher). All individual components (proteins, RNAs) and different combinations of protein:RNA complexes were purified with gel filtration in advance to be injected into SEC-MALS. 240-600 μg of sample in 100 μL volume was injected onto a Superdex 200 10/300 GL gel filtration column (GE healthcare) and at a flow rate 0.5 mL/min. Multiangle light scattering was detected by a DAWN 8+ detector (Wyatt technology). Data was analyzed with ASTRA (Wyatt technology).

2.7. Thermal shift assay (TSA)

A set of 25 mM of different buffers with 50mM NaCl was prepared in range of pH 3.00 to pH 10.25 (TSA screening buffers). 8.73 mg/mL of RocC$_{24\text{-}126}$ and 5 mg/mL of lysozyme were prepared in TSA buffer (25mM HEPES pH 7.3, 150mM NaCl, 1mM DTT). 18 μL of TSA screening buffer, 1uL of the prepared RocC or lysozyme, and 1uL of x500 sypro orange reagent (Invitrogen) was mixed and loaded in 96 well plate. Fluorescence was measured in gradual increment from 20°C to 95°C on an Applied Biosystems 7500 FAST RealTime PCR System.

2.8. Dynamic light scattering (DLS)

Polydispersity of RocC$_{24\text{-}126}$ at different salt concentration were measured by dynamic light scattering instrument using Protein solutions dynaPro (Wyatt technology). 20 μL of RocC$_{24\text{-}126}$ was prepared in 25 mM Tris pH 8.0, 10% glycerol with three different NaCl concentration, 37.5 mM, 75 mM, and 150 mM. 14 μL of samples were loaded into special cuvette for DLS and amplitude was set to be around 200. To minimize the experimental errors, each experiment was performed more than 30 photon counts. Experimental analyses were done by Dynamics program(Wyatt technology)

2.9. RocC:RocR purification for crystallization and SEC-SAXS

Both purified RocC (either RocC$_{1\text{-}126}$ or RocC$_{14\text{-}126}$) and RocR (either RocR$_{SL3}$ or RocR$_{9bp\text{-}tet}$) were in crystallization buffer (10 mM HEPES-KOH pH 7.3, 100 mM KCl, 5 mM MgCl$_2$, 1 mM DTT). RocC were incubated with an excessive RocR for 30 minutes at 4°C. The RocC:RocR complex was separated from an excessive amount of RocR by running superdex 75 16/60 (Cytiva, Figure 4-6C). RocC:RocR complex was collected and concentrated using 10k centrifugal filters for crystallization and SEC-SAXS.

2.10. Crystallization and crystallographic data collection

30 mg/mL of $\mathrm{RocC}_{24\text{-}126}$ in 25 mM Tris:HCl pH 8.0, 30 mM NaCl, 1mM DTT was crystallized by hanging drop vapor diffusion with a reservoir solution (0.2 M $NH_4CH_3CO_2$, 0.1 M HEPES pH 7.5, 10% PEG 3350) at 16°C. Protein:reservoir solution ratios of 1:1 or 1:2 yielded crystals. 32 mg/mL of $\mathrm{RocC}_{1\text{-}126}$ in 25 mM Tris:HCl pH 8.0, 30 mM NaCl, 1 mM DTT was crystallized by hanging drop vapor diffusion with a reservoir solution consisting of 0.2 M $(NH_4)_2SO_4$, 0.1 M HEPES:NaOH, pH 7.3, 25% PEG 3350 at 16°C in either ratio 1:1 or 1:2 of protein solution:resevoir solution. The quality of $\mathrm{RocC}_{1\text{-}126}$ crystals was improved using additive screens (HR2-428, Hampton research) at 4°C. 2 μL of 32 mg/mL $\mathrm{RocC}_{1\text{-}126}$ was mixed with 1.6 μL of reservoir solution (0.2 M $(NH_4)_2SO_4$, 0.1 M HEPES:NaOH, pH 7.3, 25% PEG 3350) and 0.4 μL of additive screen was mixed to make a 4 μL drop. Various additives produced high quality of crystals: multivalent ions (0.1 M Barium chloride dihydrate, 0.1M strontium chloride hexahydrate, 0.1 M Yttrium(III) chloride hexahydrate, 0.1 M Chromium (III) chloride hexahydrate), Linker (0.3 M Glycyl-glycyl-glycine), Polymer (10% w/v Polyethylene glycol 3,350), Carbohydrate (30% Sucrose, 12% w/v myo-Inositol), Organic, non-volatile solvent (30% w/v 1,6-Hexanedio), Organic, volatile solvents (40% v/v tert-Butanol, 40% v/v 1,3-Propanediol). Homogenous $\mathrm{RocC}_{14\text{-}126}$:$\mathrm{RocR}_{9\mathrm{bp\text{-}tet}}$ complex in 10 mM HEPES-KOH pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 1 mM TCEP was purified from excess free components using gel filtration chromatography and was crystallized by sitting drop vapor diffusion against a reservoir buffer containing 0.2 M Li_2SO_4, 0.1 M Tris: HCl, pH 8.5 30 % (w/v) PEG 4000 (Top96, A10, Anatrace). Crystals grew at 1:1 protein:precipitant ratio at room temperature.

Data for $\mathrm{RocC}_{24\text{-}126}$ were collected on a RIGAKU MICROMAX-007 HF with a DECTRIS PILATUS3 R200K-A detector to a final resolution of 2.10 Å. Data for $\mathrm{RocC}_{1\text{-}126}$ were collected at CLS (Canadian Light Source) BEAMLINE 08ID-1 with DECTRIS PILATUS3 S 6M detector to a resolution of 2.02 Å. Crystals of the $\mathrm{RocC}_{14\text{-}126}$:$\mathrm{RocR}_{9\mathrm{bp\text{-}tet}}$ complex were collected at ALS (Advanced Light Source) BEAMLINE 8.2.2 with ADSC QUANTUM 315r detector and data was obtained to a resolution of 3.20 Å. All data was integrated with HKL-2000 (Otwinowski & Minor, 1997).

2.11. Crystallographic structure determination and refinement

The structure of RocC$_{24\text{-}126}$ was determined by molecular replacement (MR) using Phaser (McCoy et al., 2007). A pruned model of FinO$_{85\text{-}158}$ generated using sculptor (Bunkóczi & Read, 2011) was used as a search model. 2 protomers were found in the asymmetric unit. Non-crystallographic symmetry (NCS) was used in the early stages of refinement using PHENIX (Adams et al., 2010). Missing parts of the model were manually built in Coot (Emsley et al., 2010) and refined using the PHENIX to a R_{work} of 16.1% and an R_{free} of 23.8% to a final resolution of 2.10 Å.

The structure of RocC$_{1\text{-}126}$ was solved by MR, using RocC$_{24\text{-}126}$ as a search model with Phaser(McCoy et al., 2007). Nine protomers were placed in the asymmetric unit. NCS was used throughout the refinement and revealed α-helical density for the N-terminus of RocC for two of the protomers. Manual building and refinement were carried out with Coot (Emsley et al., 2010) and PHENIX (Adams et al., 2010) with the assistance of 9-fold NCS to a final R_{work} of 19.3% and an R_{free} of 21.3% to a resolution 2.02 Å.

The RocC$_{14\text{-}126}$:RocR$_{9bp\text{-}tet}$ complex was phased by molecular replacement using the high-resolution structure of RocC$_{1\text{-}126}$ as a search model. 10 protomers were placed in the asymmetric unit and NCS was used throughout the refinement. Refinement of 10 protomers showed difference densities indicative of RNA bound to 4 of the 10 protomers. This difference density map was improved through 4-fold NCS averaging (Figure 4-9). An ideal A-form helix was then fit to the stem portion of the 4 RNA molecules. The hairpin loops were built using a related tetraloop model (PDB ID: 4Z3S). Comparison of difference maps at low vs high sigma cutoffs helped to distinguish phosphate groups from sugar and base moieties, which was particularly helpful in the building of the 3' single-stranded tail. The structure was refined using non-crystallographic symmetry restraints in PHENIX (Adams et al., 2010) to a final R_{work}/R_{free} of 21.9% and 27.1%, respectively, to a final resolution of 3.20 Å.

2.12. NMR spectroscopy

RNA samples for NMR spectroscopy were lyophilized as sodium salts and dissolved in 420 µL NMR buffer (25 mM HEPES pH 7.3, 150 mM NaCl) and transferred into standard 5mm NMR tubes giving 0.2 to 1 mM sample concentrations. All NMR experiments were recorded on Bruker 600 MHz Avance II+ NMR or Bruker 700 MHz AvanceNeo NMR spectrometers equipped with Prodigy TCI probes. The imino proton resonances in the apo state were assigned by a combination of $^1H,^1H$- jump and return NOESY experiments (150 ms mixing time, 10°C) and residue-specific $^{15}N^1$-guanosine and $^{15}N^3$-uridine labeling. The RNA protein complex was prepared by mixing the RNA with one equivalent of ^{15}N labeled RocC$_{14\text{-}126}$, followed by size exclusion chromatography. For NMR spectroscopy, the RNA protein complex was concentrated to a volume of 420 μL by ultracentrifugation (molecular weight cut off 3 kDa).

2.13. Searching the PDB for examples of protein-RNA interaction motifs observed in RocC-RocR

The duplex portion of the RocR terminator is recognized by an N-capped helix (N-cap motif) in RocC. To search for similar interactions in the PDB, we wrote a Perl script that could identify an N-capping Ser/Thr residue which, together with two C-terminal mainchain NH groups, H-bond with two consecutive phosphate groups (see Ncap_RNA.pl, supplemental files). We searched the entire protein – nucleic acid structure database (all X-ray and cryoEM structures as of April, 2021 – 8935 structures). We visualized the hits in Pymol, aligned on the nucleotides bound by the N-cap. In this way, we uncovered a similar binding interaction between an N-cap motif in the ROQ domain of the T-cell regulatory protein, Roquin, with a region of duplex RNA within a target UTR (Figure 4-11D).

The 3' nucleotide of RocR is recognized by a conserved pocket in RocC that contains a β-hairpin-helix motif composed of amino acids 50-54. To search for similar nucleotide binding motifs in other structures, we wrote a Perl script that could identify structures that have consecutive backbone NHs that are within hydrogen bonding distance to the O2' and O3' atoms of a nucleotide (see 3pocket.pl, supplemental files). We used this script to scan all protein-RNA and protein-nucleotide structures determined by either X-ray crystallography or cryoEM that were available in the PDB (as of April 2021). We then visualized the hits aligned with the RocC-RocR structure in Pymol and identified an interaction in the eukaryotic ribosomal small subunit as most similar to RocC-RocR. The interaction involves a nucleotide within the 18S rRNA and a β-hairpin-helix motif within the 40S ribosomal protein S24E (Figure 4-11B).

2.14. Size-exclusion chromatography with small-angle X-ray scattering (SEC-SAXS)

SEC-SAXS data of RocC$_{14\text{-}126}$:RocR$_{9bp\text{-}tet}$, RocC$_{1\text{-}126}$:RocR$_{SL3}$, RocC$_{1\text{-}126}$, was collected at beamline 12.3.1 at the Advanced Light Source (ALS). The 19 mg/mL of complex was in the crystallization buffer (10 mM HEPES-KOH pH 7.5, 100 mM KCl, 5 mM MgCl2, 1 mM TCEP, all buffer prepared in DEPC treated water) was loaded onto a PROTEIN KW-802.5 column (Shodex). Sample was ran at 0.5 mL/min in the crystallization buffer and scattering was collected using a PILATUS3 X 2M detector. Scattering data of RocC$_{1\text{-}126}$:RocR$_{SL3}$ from asymmetric eluent peak was deconvoluted using RAW to separate compact and extended scattering profiles (Hopkins et al., 2017).

The disconnected part of RocC$_{1\text{-}126}$ was connected using SWISS-MODEL (Waterhouse et al., 2018) and fit into RocC$_{1\text{-}126}$ SAXS envelope using Multi-FoXS (Schneidman-Duhovny et al., 2016). RocC$_{SL3}$ was modeled based on crystal structure of RocC$_{9bp\text{-}tet}$ and 3D prediction of RocC$_{SL3}$ from RNAcomposer. Stem of the SL3 and 3' tail having direct contact with RocC in the crystal structure and upper stem and 7 nucleotide loop from RNA composer was manually combined using Pymol (Schrödinger). RocC$_{1\text{-}126}$:RocR$_{SL3}$ model was modeled using FoXS restraint by deconvoluted "compact profile" and "extended profile". Solution model of RocC$_{14\text{-}126}$:RocR$_{9bp\text{-}tet}$ was generated using crystal structure of RocC$_{14\text{-}126}$:RocR$_{9bp\text{-}tet}$ restrainted by scattering profile without deconvolution using FoXS.

2.15. Fluorescence polarization assay (FP)

5' FAM-labeled RocR$_{SL3}$ was purchased from IDT and 5' FAM-labeled RocR$_{3nt}$, RocR$_{8nt}$ were chemically synthesized. 5 μL of 80 nM RNA in FP reaction buffer (25 mM HEPES pH 7.3, 150 mM NaCl, 4 mM MgCl$_2$, 10% glycerol, 1 mM DTT, 0.4 mg/mL yeast tRNA (ThermoFisher)) was mixed with 15 μL of point mutated RocC$_{14\text{-}126}$ to a final protein concentration of between 320 μM and 2.4 nM in a 20 μL reaction. Reactions were transferred to a 384-well plate and were incubated for 1.5 hours at 25°C. FP experiments were conducted an Envision 2103 multilabel plate reader (Perkin Elmer) using 485 nm excitation and recorded at 538 nm. All experiments were repeated three times and the dissociation constant was calculated by fitting results to a sigmoidal curve with a 4-parameter logistic (4PL) equation:

$$y = d + \frac{(a-d)}{(1+\frac{x}{c})^b}$$

Where a and d are the estimated minimum and maximum values. b is the slope factor and c is c_{50}.

2.16. Isothermal calorimetry assay (ITC)

ITC was performed using MicroCal PEAQ-ITC (Malvern panalytical) at 25°C. 10 nucleotide RNA (5'-GGCCCUUUCU-3') was purchased from IDT and SL3-9bp-tet was chemically synthesized. $\text{RocC}_{14\text{-}126}$ was prepared as described above. All components were dialyzed in ITC buffer (25 mM HEPES pH 7.3, 150 mM NaCl, 10% glycerol, 1 mM TCEP). 24-50 μM of protein in the sample cell was titrated with RNA titrand at concentrations between 200-400 μM. 2 μL (except for first injection) of concentrated titrand were injected 19 times every 180-240 seconds. Detailed individual titration condition is on Table 2-4. Obtained data were analyzed by MicroCal PEAQ-ITC Analysis Software (Malvern panalytical).

2.17. Bacterial strains and growth conditions

The *L. pneumophila* strains in this study are derived from the Paris clinical isolate (Outbreak isolate CIP107629). These strains (genotypes in Table 2-5 and construction details in Supplementary Material and Methods) were grown in liquid media ACES [N-(2-acetamido)-2-aminoethanesulfonic acid]-buffered yeast extract (AYE) or on solid media ACES-buffered

Table 2-4. ITC experimental conditions.

	Sample cell	Syringe	Cell concentration (μM)	Syringe concentration (μM)	Interval (sec)
RocC:10nt RNA	Protein	RNA	50	400	240
ProQ:10nt RNA	Protein	RNA	30	400	240
FinO:10nt RNA	Protein	RNA	50	400	240
RocC:$\text{RocR}_{\text{SL3-9bp-tet}}$	RNA	Protein	30	350	240
ProQ:$\text{RocR}_{\text{SL3-9bp-tet}}$	Protein	RNA	40	394	240
FinO:$\text{RocR}_{\text{SL3-9bp-tet}}$	RNA	Protein	24	200	180

Table 2-5. Bacterial strains, plasmids and oligonucleotides used in this study.

Strains	Relevant genotype	Reference
L. pneumophila		
Paris WT	Paris Outbreak isolate CIP107629	CNR Lyon
Paris *rocC*$_{TAA}$	Paris; *rocC*$_{TAA}$ (previously noted *lpp0148*$_{TAA}$)	(Juan, Attaiech, and Charpentier 2015)
Paris Δ*rocC::MK*	Paris; *rocC::(lacIq*, P$_{tac}$-*mazF*, *nptII*); KanR, IPTG^S	(Juan, Attaiech, and Charpentier 2015)
Paris *ΔrocR*	Paris; *ΔrocR*	(Attaiech et al. 2016)
Paris *rocC, kan*	Paris transformed by pLFP01; *rocC*_*lpp0149*::(*nptII*); KanR	This study
Paris *rocC*$_{K15D}$, *kan*	Paris transformed by pLLA96; *rocC*$_{K15D}$_*lpp0149*::(*nptII*); KanR	This study
Paris *rocC*$_{Q17A}$, *kan*	Paris transformed by pLLA112; *rocC*$_{Q17A}$_*lpp0149*::(*nptII*); KanR	This study
Paris *rocC*$_{K18D}$, *kan*	Paris transformed by pLLA97; *rocC*$_{K18D}$_*lpp0149*::(*nptII*); KanR	This study
Paris *rocC*$_{S21A}$, *kan*	Paris transformed by pLLA113; *rocC*$_{S21A}$_*lpp0149*::(*nptII*); KanR	This study
Paris *rocC*$_{I51A}$, *kan*	Paris transformed by pLLA98; *rocC*$_{I51A}$_*lpp0149*::(*nptII*); KanR	This study
Paris *rocC*$_{S70A}$, *kan*	Paris transformed by pLLA99; *rocC*$_{S70A}$_*lpp0149*::(*nptII*); KanR	This study
Paris *rocC*$_{S72A}$, *kan*	Paris transformed by pLLA111; *rocC*$_{S72A}$_*lpp0149*::(*nptII*); KanR	This study
Paris *rocC*$_{S70A\text{-}S72A}$, *kan*	Paris transformed by pLLA106; *rocC*$_{S70A\text{-}S72A}$_*lpp0149*::(*nptII*); KanR	This study
Paris *rocC*$_{K71D}$, *kan*	Paris transformed by pLLA100; *rocC*$_{K71D}$_*lpp0149*::(*nptII*); KanR	This study
Paris *rocC*$_{K73D}$, *kan*	Paris transformed by pLLA101; *rocC*$_{K73D}$_*lpp0149*::(*nptII*); KanR	This study
Paris *rocC*$_{R75D}$, *kan*	Paris transformed by pLLA102; *rocC*$_{R75D}$_*lpp0149*::(*nptII*); KanR	This study
Paris *rocC*$_{T82A}$, *kan*	Paris transformed by pLLA103; *rocC*$_{T82A}$_*lpp0149*::(*nptII*); KanR	This study
Paris *rocC*$_{R83D}$, *kan*	Paris transformed by pLLA104; *rocC*$_{R83D}$_*lpp0149*::(*nptII*); KanR	This study
Paris *rocC*$_{Y87F}$, *kan*	Paris transformed by pLFP08; *rocC*$_{Y87F}$_*lpp0149*::(*nptII*); KanR	This study
Paris *rocC*$_{R97M}$, *kan*	Paris transformed by pLLA109; *rocC*$_{R97M}$_*lpp0149*::(*nptII*); KanR	This study
Paris *rocC*$_{N115A}$, *kan*	Paris transformed by pLLA110; *rocC*$_{N115A}$_*lpp0149*::(*nptII*); KanR	This study
Paris *rocC*$_{K119D}$, *kan*	Paris transformed by pLLA105; *rocC*$_{K119D}$_*lpp0149*::(*nptII*); KanR	This study
Paris *rocCΔN14*	Paris Δ*rocC::MK* transformed by PCR DE$_1$; *rocCΔN14*	This study
Paris *rocCΔN19*	Paris Δ*rocC::MK* transformed by PCR DE$_3$; *rocCΔN19*	This study
Paris *rocCΔN24*	Paris Δ*rocC::MK* transformed by PCR DE$_5$; *rocCΔN24*	This study
E. coli		
DH5α, λpir	F$^-$, *sup*E44, Δ*lac*U169 (Φ*lacZ*ΔM15), *recA*1, *endA*1, *hsdR*17, *thi-1*, *gyrA*96, *relA*1, λ*pir* lysogen	Laboratory strain collection
BL21-Gold (DE3)	*F$^-$ ompT hsdS(r$_B^-$ m$_B^-$) dcm$^+$ Tet^r gal λ(DE3) endA Hte*	Laboratory strain collection

Plasmids	Relevant genotype	Reference
pET-47b(+) RocC$_{1\text{-}230}$	pET-47b(+) derivative; IPTG-inducible production of His-(HRV-3C)-Lpp0148$_{1\text{-}230}$ (Paris); KanR	(Attaiech et al. 2016)
pGEM-ihfB::Kan	pGEM-T Easy; *ihfB*::*nptII*; AmpR, KanR	(Juan, Attaiech, and Charpentier 2015)
pGEM-MK	pGEM-T Easy; FRT-(*lacIq*, P$_{tac}$-*mazF*, *nptII*)-FRT; AmpR, KanR	(Bailo et al. 2019)
pGEMPKD4	pGEM-T Easy; FRT- *nptII* -FRT; AmpR, KanR	(Attaiech et al. 2016)
pGEX-6P-1	pBR322 derivative; IPTG-inducible production of GST-(HRV-3C)-protein; used for in vitro production of RocC mutants; AmpR	Cytiva 28-9546-48
pLFP01	pGEM-T Easy; *rocC_lpp0149*::(*nptII*); AmpR, KanR	This study
pLFP08	pLFP01 derivative; *rocC*$_{Y87F}$*_lpp0149*::(*nptII*); AmpR, KanR	This study
pLLA96	pLFP01 derivative; *rocC*$_{K15D}$*_lpp0149*::(*nptII*); AmpR, KanR	This study
pLLA97	pLFP01 derivative; *rocC*$_{K18D}$*_lpp0149*::(*nptII*); AmpR, KanR	This study
pLLA98	pLFP01 derivative; *rocC*$_{I51A}$*_lpp0149*::(*nptII*); AmpR, KanR	This study
pLLA99	pLFP01 derivative; *rocC*$_{S70A}$*_lpp0149*::(*nptII*); AmpR, KanR	This study
pLLA100	pLFP01 derivative; *rocC*$_{K71D}$*_lpp0149*::(*nptII*); AmpR, KanR	This study
pLLA101	pLFP01 derivative; *rocC*$_{K73D}$*_lpp0149*::(*nptII*); AmpR, KanR	This study
pLLA102	pLFP01 derivative; *rocC*$_{K75D}$*_lpp0149*::(*nptII*); AmpR, KanR	This study
pLLA103	pLFP01 derivative; *rocC*$_{T82A}$*_lpp0149*::(*nptII*); AmpR, KanR	This study
pLLA104	pLFP01 derivative; *rocC*$_{R83D}$*_lpp0149*::(*nptII*); AmpR, KanR	This study
pLLA105	pLFP01 derivative; *rocC*$_{K119D}$*_lpp0149*::(*nptII*); AmpR, KanR	This study
pLLA106	pLFP01 derivative; *rocC*$_{S70A\text{-}S72A}$*_lpp0149*::(*nptII*); AmpR, KanR	This study
pLLA109	pLFP01 derivative; *rocC*$_{R97M}$*_lpp0149*::(*nptII*); AmpR, KanR	This study
pLLA110	pLFP01 derivative; *rocC*$_{N115A}$*_lpp0149*::(*nptII*); AmpR, KanR	This study
pLLA111	pLFP01 derivative; *rocC*$_{S72A}$*_lpp0149*::(*nptII*); AmpR, KanR	This study
pLLA112	pLFP01 derivative; *rocC*$_{Q17A}$*_lpp0149*::(*nptII*); AmpR, KanR	This study
pLLA113	pLFP01 derivative; *rocC*$_{S21A}$*_lpp0149*::(*nptII*); AmpR, KanR	This study

Cloning primers	Sequence (5’ to 3’)
lpp0148_P1	GCCGTTTTAAATCGGCCAGAAAG
lpp0148_P2	ggcccaattcgccctatagtgagtcgCAGGCAAGCCAGATAATCAAGAC
lpp0148_P3	gggtttgctcgggtcggtggcataTGAAAGCTCGTGAAGTCCGTATCG
lpp0148_P4	CTTCCCCCTAAAAATCAGGATGTC
lpp0148RM_P3	GAACTAAGGAGGATATTCATATGGACCATGGCTGTTTATTTCTCCATTATGGGCTG
lpp0148RM_R	GGAACTTCGAAGCAGCTCCAGCCTACACAATCTGCTGAAAATAATGTCGCTATTCGC
lpp0149pX3_R	ATCACCCGGGCTAAGGTTGTTTTAAAATCATAAAGTAAGC
MazFk7-F	CGACTCACTATAGGGCGAATTGGGCCGCTTTCCAGTCGGGAAACCTG
MazF-R	CATATGCCACCGACCCGAGCAAACCCGAAGAAGTTGTCCATATTGGCCAC
pKD4_P1	GATTGTGTAGGCTGGAGCTGCTTCG
pKD4_P2	GCCATGGTCCATATGAATATCCTCC

charcoal yeast extract (CYE) plates at 30°C or 37°C. Liquid cultures were performed in 13-mL tube containing 3 mL of medium in a shaking incubator at 200 rpm. When appropriate, kanamycin and streptomycin were used at 15 µg/mL or 50 µg/mL, respectively.

Escherichia coli strains (Table 2-5) were cultivated in LB medium with shaking or on LB-agar plates at 37°C. When appropriate, kanamycin and ampicillin were used at 50 µg/mL and 100 µg/mL, respectively.

2.18. Construction of plasmids and strains

Construction of L. pneumophila strains with mutated RocC, punctual mutants

We first created the pLFP01 plasmid. Using PCR (PrimeStarMax, Takara), we amplified *rocC* from the Paris WT strain using primer pair lpp0148_P1/lpp0148RM_R, a kanamycin resistance cassette (*nptII*) from the pGEMPKD4 using primer pair pKD4_P1/pKD4_P2 and lpp0149 (which follows *rocC* on the genome) from the Paris WT strain using primer pair lpp0148RM_P3/lpp0149pX3_R. The 3 PCR fragments were assembled by PCR overlap extension using primer pair lpp0148_P1/lpp0149pX3_R and cloned into the pGEM-T Easy vector (Promega). After transformation in *E. coli* DH5α, λpir, transformants were selected on LB plates containing kanamycin. The plasmid was verified by PCR and sequencing.

The plasmid pLFP01 [*rocC-lpp0149*::(*nptII*)] was submitted to site-directed mutagenesis to create the different punctual mutants of RocC. PCRs were done using primer pairs designed to change the desired amino acid (Table 2-6) on pLFP01 as template. PCR products were then digested by DpnI to remove the parental pLFP01 and transformed in *E. coli* DH5α λpir. Transformants were selected on LB plates with kanamycin. Plasmids were verified by PCR and sequencing.

Each plasmid was then used to transform *L. pneumophila* Paris *rocC*$_{TAA}$ by natural transformation. As these plasmids are non-replicative in *L. pneumophila*, the internalized molecules recombine with the chromosome via a double cross-over allowing the replacement of the *rocC*$_{TAA}$ allele with the mutated *rocC* allele. Transformants were selected on CYE plates with kanamycin. The *rocC* locus was verified by PCR and sequencing, and the presence of the protein was verified by Western-Blot. For each mutant, 2 independent clones were kept and tested.

Table 2-6. Oligonucleotides used to create pLFP01 derivatives with punctual RocC mutants.

Mutation	**Forward primer**		**Obtained plasmid**
	name	**sequence (5'-3')** (mutations are in lowercase)	
K15D	LA161_rocCk15d_F	GAACCGCTGTCATCAATGACGCACAAAAAAATCAATCC	pLLA96
Q17A	LA189_rocCq17a_F	CCGCTGTCATCAATAAAGCAGCAAAAAATCAATCCAAGCGCG	pLLA112
K18D	LA163_rocCk18d_F	CATCAATAAAGCACAAGACAATCAATCCAAGCGCG	pLLA97
S21A	LA191_rocCs21a_F	CAATAAAGCACAAAAAAATCAAGCCAAGCGCGCGCGATCTG	pLLA113
I51A	LA165_rocCi51a_F	GTATTCGGCCATTAAAGGCTGGTATTATGTCTGATATATTG	pLLA98
S70A	LA167_rocCs70a_F	GCAGAGCAAGTTGGAGTTGCTAAAAGCAAATTAAGGGAAGC	pLLA99
S72A	LA187_rocCs72a_F	GCAAGTTGGAGTTTCTAAAGCCAAATTAAGGGAAGCTGTTG	pLLA111
S70A, S72A	LA181_rocCs70.72a_F	GCAGAGCAAGTTGGAGTTGCTAAAGCCAAATTAAGGGAAGCTGTTG	pLLA106
K71D	LA169_rocCk71d_F	GAGCAAGTTGGAGTTTCTGACAGCAAATTAAGGGAAGCTG	pLLA100
K73D	LA171_rocCk73d_F	GTTGGAGTTTCTAAAAGCGACTTAAGGGAAGCTGTTGTTC	pLLA101
R75D	LA173_rocCr75d_F	GAGTTTCTAAAAGCAAATTAGACGAAGCTGTTGTTCTTTTTAC	pLLA102
T82A	LA175_rocCt82a_F	GGAAGCTGTTGTTCTTTTTGCCCGTCGTCTTGATTATCTG	pLLA103
R83D	LA177_rocCr83d_F	GCTGTTGTTCTTTTTACCGATCGTCTTGATTATCTGGCTTG	pLLA104
Y87F	FP7_paris-rocC35_F	TTTTACCCGTCGTCTTGATTTTCTGGCTTGCCTGAAAGCTC	pLFP08
R97M	LA183_rocCr97m_F	TGAAAGCTCGTGAAGTCATGATCGATTTGCATGGAAATCC	pLLA109
N115A	LA185_rocCn115a_F	CTGAGGAAGAAGCGGAGGCTGCTTCCATGAAAATTAAAAAACGCG	pLLA110
K119D	LA179_rocCk119d_F	GAAGCGGAGAATGCTTCCATGGACATTAAAAAACGCGTGGAAAAG	pLLA105

Mutation	**Reverse primer**		**Obtained plasmid**
	name	**sequence (5'-3')** (mutations are in lowercase)	
K15D	LA162_rocCk15d_R	GATTTTTTTGTGCGTCATTGATGACAGCGGTTC	pLLA96
Q17A	LA190_rocCq17a_R	CGCGCTTGGATTGATTTTTTGCTGCTTTATTGATGACAGCGG	pLLA112
K18D	LA164_rocCk18d_R	CGCGCTTGGATTGATTGTCTTGTGCTTTATTGATG	pLLA97
S21A	LA192_rocCs21a_R	CAGATCGCGCGCGCTTGGCTTGATTTTTTTGTGCTTTATTG	pLLA113
I51A	LA166_rocCi51a_R	CAATATATCAGACATAATACCAGCCTTTAATGGCCGAATAC	pLLA98
S70A	LA168_rocCs70a_R	GCTTCCCTTAATTTGCTTTTAGCAACTCCAACTTGCTCTGC	pLLA99
S72A	LA188_rocCs72a_R	CAACAGCTTCCCTTAATTTGGCTTTAGAAACTCCAACTTGC	pLLA111
S70A, S72A	LA182_rocCs70.72a_R	CAACAGCTTCCCTTAATTTGGCTTTAGCAACTCCAACTTGCTCTGC	pLLA106
K71D	LA170_rocCk71d_R	CAGCTTCCCTTAATTTGCTGTCAGAAACTCCAACTTGCTCTG	pLLA100
K73D	LA172_rocCk73d_R	GAACAACAGCTTCCCTTAAGTCGCTTTTAGAAACTCCAACTTG	pLLA101
R75D	LA174_rocCr75d_R	GTAAAAAGAACAACAGCTTCGTCTAATTTGCTTTTAGAAACTC	pLLA102
T82A	LA176_rocCt82a_R	CAGATAATCAAGACGACGGGCAAAAAGAACAACAGCTTCC	pLLA103
R83D	LA178_rocCr83d_R	CAAGCCAGATAATCAAGACGATCGGTAAAAAGAACAACAGC	pLLA104
Y87F	FP7_paris-rocC35_R	GAGCTTTCAGGCAAGCCAGAAAATCAAGACGACGGGTAAAA	pLFP08
R97M	LA184_rocCr97m_R	GATTTCCATGCAAATCGATCATGACTTCACGAGCTTTC	pLLA109
N115A	LA186_rocCn115a_R	CGTTTTTTAATTTTCATGGAAGCAGCCTCCGCTTCTTCCTCAG	pLLA110
K119D	LA180_rocCk119d_R	CTTTTCCACGCGTTTTTTAATGTCCATGGAAGCATTCTCCGCTTC	pLLA105

Construction of L. pneumophila strains with mutated RocC, Nterminal deletions

Markerless Nterminal deletions of *rocC* were constructed in two steps, taking advantage of the counterselectable MK cassette. This MK cassette bears a kanamycin resistance gene and the toxin-encoding *mazF* gene under the control of an IPTG-inducible promoter (*lacIq*, P_{tac}-*mazF*, *nptII*) and was cloned into the pGEM-T Easy vector (Promega) to create plasmid pGEM-MK (Bailo et al., 2019). Insertion of MK in *rocC* can be selected on CYE plates with kanamycin after transformation of a composite PCR "upstream *rocC* - MK - downstream *rocC"*. Replacement of the *rocC*::MK allele by the *rocC*ΔNter alleles can be selected on CYE+IPTG after transformation of a composite PCR (upstream - *rocC*ΔNter – downstream). To make the Paris *rocC*::MK strain, the upstream (PCR A: lpp0148_P1 / lpp0148_P2, 2050 bp) and downstream (PCR C: lpp0148_P3 / lpp0148_P4, 2015 bp) regions of *rocC* were amplified from Paris WT chromosome with primers carrying 30-nt sequences complementary to the ends of the MK cassette (PCR B, 3177 bp amplified from plasmid pGEM-MK with primer pair MazFk7-F/MazF-R). The 3 fragments were joined by PCR overlap extension (PCR lpp0148_P1 / lpp0148_P4, 7190 bp) and used to transform *L. pneumophila* WT by natural transformation. Transformants were selected on CYE plates with kanamycin and tested for sensitivity to IPTG. Integration of the MK cassette at the correct locus was verified by PCR.

To obtain the markerless *rocC*ΔNter mutants, a second step was performed as follows. The upstream and *rocC* regions were amplified with primers carrying a 30 nucleotide tail sequence designed to create different *rocC*ΔNter alleles (PCR D and PCR E, details). These PCRs were assembled by PCR overlap extension using primer pair lpp0148_P1 / lpp0148_P4 (4 kb) and used to transform the Δ*rocC*::MK strain. Transformants were selected on CYE plates with IPTG and tested for sensitivity to kanamycin. Proper replacement of the *rocC* allele was verified by PCR and sequencing, and the presence of the protein was verified by Western-Blot. For each mutant, 2 independent clones were kept and tested.

2.19. Transformability assay

Transformability of *L. pneumophila* strains was assessed as previously described using conditions in which the WT strain is not transformable but transformation is highly efficient in hypercompetent mutant (such as the *rocC$_{TAA}$* strain which lacks RocC)(Attaiech et al., 2016).

The strains were streaked on CYE solid medium from the -80°C frozen stock culture and incubated 3 days at 37°C. The strains were then re-streaked on a new CYE plate and incubated overnight at 37°C to obtain exponentially growing cells. Bacteria were resuspended to an OD_{600}=1 (≈10^9 cells/mL) in 3 mL AYE. Two times 1 mL of cell suspension were centrifuged for 3 minutes at 5000 g in a table-top microcentrifuge, and each pellet was resuspended in 50 µL of AYE with or without transforming DNA. Each suspension was spotted on a new CYE plate and let to dry. The plate was incubated overnight at 37°C. Each spot was resuspended in 200 µL AYE. Ten-fold serial dilutions were then plated on non-selective medium and selective medium. Plates were incubated for 72 hours at 37°C and colony-forming units (CFU) counting was performed. Transformation frequency is the ratio of the number of CFUs counted on selective medium divided by the number of CFUs counted on non-selective medium. To test the transformability of the RocC punctual mutants (stains KanR), the DNA used was 2 µg (for 10^9 cells) of a 4-kb PCR fragment centered on a mutated allele of the *rpsL* gene. Transformants were thus selected on CYE+Streptomycin. To test the transformability of the RocCΔN mutants, the DNA used was either the same *rpsL*^{R} PCR as above or 2 µg (for 10^9 cells) of the pGEM-ihfB::Kan, a plasmid containing a kanamycin-resistance cassette (KanR) inserted in the *ihfB* gene of *L. pneumophila*. As this plasmid is non-replicative in *L. pneumophila*, the internalized molecules recombine with the chromosome via a double crossover allowing the integration of the KanR cassette in the *ihfB* locus. Transformants were thus selected on CYE+Streptomycin or CYE+Kanamycin. For each tested mutants, transformation assays were performed 3 to 7 times.

2.20. Western blot analysis of RocC and RocC mutants

Cells were grown at 37°C in AYE+Kan and were harvested at an OD_{600} of approximately 1 by centrifugation at 5000 g for 5 min. The pellet was resuspended directly in 125 mM Tris-HCl pH 6.8; 2% SDS; 15% glycerol; 0,005% Bromophenol Blue; 100 mM β-mercaptoetanol and incubated for 5 minutes at 85°C. Samples were then separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). Next, proteins were transferred to a polyvinylidene difluoride

(PVDF) membrane on a Trans-Blot Turbo Transfer System (Bio-Rad). RocC and RocC mutants were detected with polyclonal anti-RocC antibodies and anti-rabbit IgG, HRP conjugated antibodies (Sigma) and revelation was done using the ECL system (Thermo Scientific) according to the manufacturer's instructions. Luminescence signals were acquired using an imaging workstation equipped with a charge-coupled device camera (Thermo Scientific).

Chapter 3.

The FinO domain of RocC recognizes the base of hairpin and 3’ tail length of rho-independent transcriptional terminators

3.1. Overview

Co-immunoprecipitation coupled with RNA-deep sequencing experiments revealed that the FinO family protein RocC selectively binds to RocR to regulate competence-associated regulons. The FinO domain of RocC binds through the Rho-independent terminator SL3 of RocR and protects RocR from RNase degradation. However, how RocC specifically recognizes RocR is unknown. We characterize the stoichiometry of binding of the RocC N-terminal FinO domain with the SL3 of RocR RNA. Various RocC N-terminal truncation mutants containing the FinO domain are examined to find a minimal construct which tightly binds SL3. SL3 derivatives, which vary in the size of the loop, length of the stem and tail, are explored to understand determinants recognized by RocC. Also, binding studies of point mutations in SL3 allowed us to assess if RocC binds RNA in a sequence-specific manner.

3.2. Introduction

Regulatory RNAs are involved in many aspects of cellular processes and often do so with the assistance of RNA binding protein (RBP) partners (Hör et al., 2018; Watkins & Arya, 2019). The FinO family of proteins is becoming widely acknowledged as an important group of RNA chaperones in bacteria (Mark Glover et al., 2015). The chaperone activity of FinO begins with the recognition of the *traJ* mRNA 5' UTR and its antisense RNA, FinP, by the FinO domain of the FinO protein (Biesen & Frost, 1994). While the FinO domain binds to the Rho-independent terminator structures of SLII of FinP and SLIIc of *traJ* mRNA, the flexible extension of the FinO N-terminus promotes FinP-*traJ* UTR hybridization, initiated with a kissing interaction involving the hairpin loops within the complementary regions of the RNAs. The degradation of the resulting heteroduplex by RNase III leads to the repression of bacterial conjugation (Jerome et al., 1999). This fine control begins with specific recognition of the sRNA, and the detailed RNA recognition mechanisms by the FinO domain are still under investigation.

FinP was initially predicted to consist of two stem-loops connected with a short linker (Finlay, B. B. et al., 1986; Koraimann et al., 1991). Partial RNA digestion using single- and double-stranded nucleases mapped FinP structures and this mapping empirically proved the predicted structure (van Biesen et al., 1993). A series of EMSAs with radiolabeled RNA showed that FinO mainly interacts through SLII and required the stem as well as 5' 4 nucleotide and 3' 6 nucleotide extensions for tight binding (Jerome & Frost, 1999). Sequence alignment of SLIIs from different plasmids showed diverse sequences in loop (Finlay, B. B. et al., 1986), suggesting that the loop is dispensable for FinO recognition (Jerome & Frost, 1999). Footprinting assays of FinO bound to SLII and SLIIc revealed that FinO binds the base of the stem and 3' and 5' tails of its target RNAs (Arthur et al., 2011). Furthermore, the chemical structure of the 3’ ribose sugar was shown to be critical for binding. Either 3’ phosphorylation, or oxidation of the 3’ nucleotide were shown to abrogate FinO binding to SLII (Arthur et al., 2011). A recent study found a new target of FinO in *S. enterica*, RepX, which contains a similar Rho-independent terminator (El Mouali et al., 2021).

Development of extensive parallel RNA sequencing methods revealed new targetomes of the FinO family. Grad-seq revealed ProQ in *S. enterica* as an RNA binding hub, interacting with a number of structured RNAs (Smirnov et al., 2016). UV crosslinking and immunoprecipitation followed by RNA-sequencing in *E. coli*, *S. enterica*, and *N. meningitidis* confirmed that ProQ proteins in these bacteria bind a range of RNAs and likely rely on direct interactions with the 3’ intrinsic transcriptional terminator RNA (Bauriedl et al., 2020; Holmqvist et al., 2018). Stein and colleagues showed that *E. coli* ProQ binds to terminators with at least a 2 base-pair stem and a 4 nucleotide 3’ tail (Stein et al., 2020).

The RNA binding behaviours of the FinO family all seem to involve recognition of a 3’ transcriptional terminator structure, however in some cases, the binding is highly specific for a single terminator (eg: FinO) whereas in other cases, binding can be more promiscuous (eg: ProQ). Even though the understanding of the range of FinO family targets has expanded over the last five years, the molecular mechanisms by which these proteins interact with their targets is still largely unknown. To achieve this goal, we set out to study a highly specific member of the FinO family: the *Legionella* protein RocC and its RNA target RocR. We designed and tested minimal RocC constructs that bind to the known target, stem-loop 3 (SL3) of RocR. Determinant searches on SL3 were initiated by dissecting SL3 into three elements: the 7 nucleotide loop, the 11 base-pair stem, and the 5 nucleotide 3’ tail. By changing the sizes of each component, we found that RocC requires at a minimum, a 5 base-pair stem, and has a high degree of selectivity for 5 nucleotide 3’ tail length. We tested the sequence specificity within 5 nucleotide poly pyrimidine tail through the examination of a variety of substitution mutants. We did not observe any dramatic binding affinity changes by substituting poly(U) tail components, suggesting that RocC does not bind the tail in a sequence-dependent manner. Together, our work revealed that $\text{RocC}_{14\text{-}126}$ is a minimal FinO domain-containing protein that forms a 1:1 complex with an intrinsic terminator, consisting of at least a 5 base-pair stem and specifically a 5 nucleotide 3' polypyrimidine tail.

3.3. Results

3.3.1. Defining a minimal RNA binding domain of RocC

Preliminary EMSA demonstrated that the isolated FinO domain of RocC interacts with its binding partner RocR in the low μM range through interactions with the SL3 terminator hairpin (Attaiech et al., 2016). To obtain a structural understanding of RocC, we modeled the 3D structure of RocC full length using the SWISS-MODEL homology modelling program. This program automatically chose the *E. coli* ProQ NMR structure as a template which has 34% sequence identity to the $\text{RocC}_{16\text{-}129}$ structure (Figure 3-1A). Regions that potentially form secondary structures were predicted with relatively high confidence in QMEANDisCo (Figure 3-1C) (Studer et al., 2020). QMEANDisCo scores the similarities among a large set of predicted models. Well-conserved regions will show less divergence between the predicted models, resulting in higher scores in the prediction. Unstructured regions are difficult to model consistently and they generally result in low confidence predictions. The predicted model also showed a reasonably similar structure to FinO (RMSD 2.7 Å^2) (Figure 3-1B). SWISS-MODEL failed to model the extreme N-terminus (residues 1-15) and around 100 residues in C-terminus (130-230) since it could not find a homology model to predict this part of the protein. Overlay of all existing FinO family structures as of 2016 showed that $\text{RocC}_{36\text{-}117}$ was structurally conserved with ProQ and FinO (Figure 3-1B).

We carried out limited proteolysis to search for a minimal, stably folded core of RocC, which still binds to SL3. Limited proteolysis is a powerful technique that is prevalently used to search for stable, structural domains for biochemical analysis and which are often excellent targets for crystallization. *E. coli* FinO, which was the first determined structure among the FinO family, is a good example. Limited proteolysis provided a list of proteolytically stable constructs, and $\text{FinO}_{26\text{-}186}$, which structure was determined, was one of these candidates (Ghetu et al., 1999). Limited proteolysis of full-length RocC revealed the existence of a proteolytically-stable core, roughly half size of intact RocC (Figure 3-2). Peptide cutter, a proteolytic degradation prediction program, predicted cutting sites for trypsin, which cleaves C-terminal to arginine and lysine, and

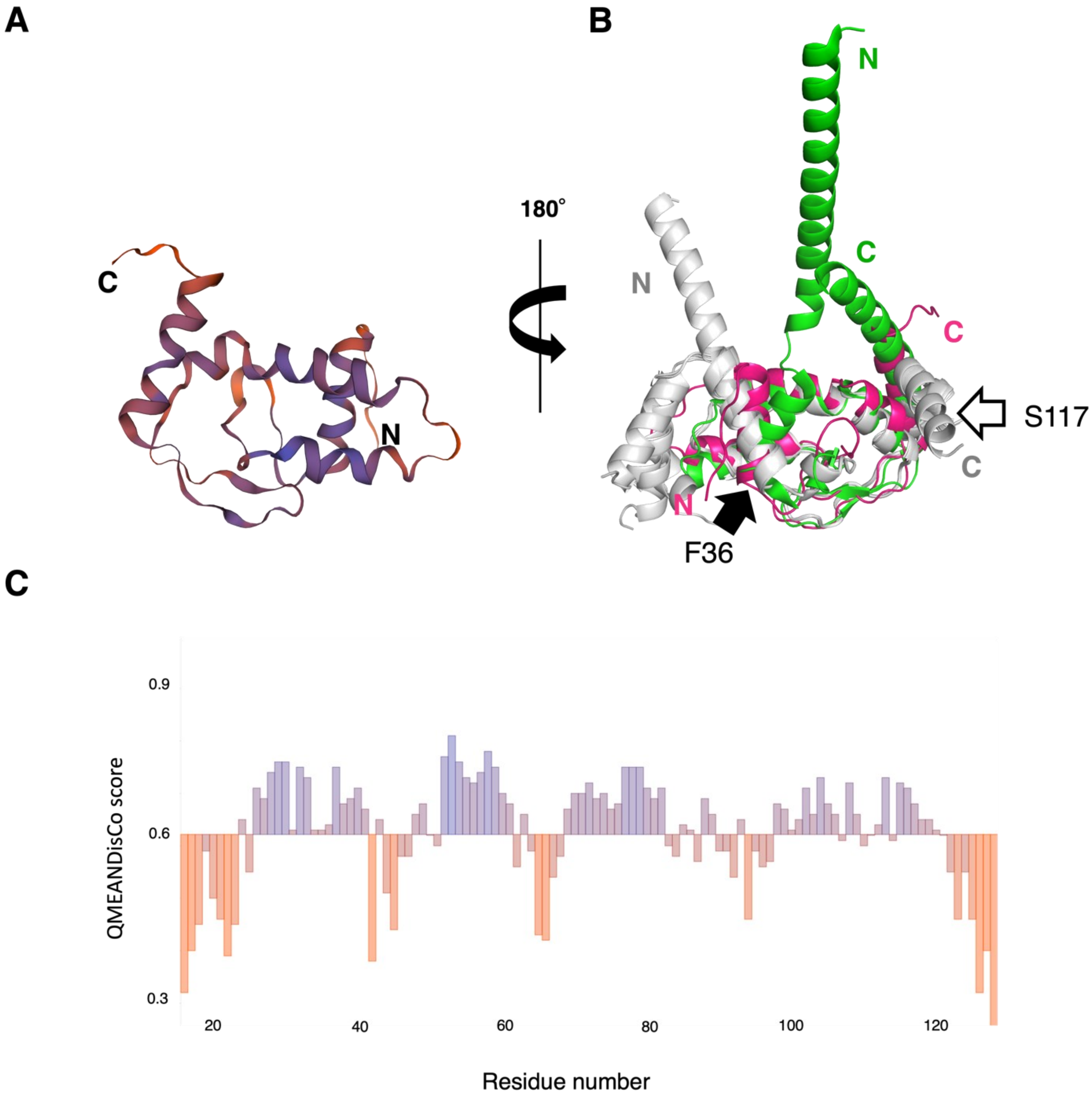


Figure 3-1. Predicted FinO domain of RocC and overlay of the RocC FinO domain with existing crystal structures.

(A) Predicted FinO domain of RocC. (B) Overlay view of the predicted FinO domain of RocC (magenta), six copies of NMB 1681 crystal structure (light grey), and FinO (green). Two arrows indicate where the structures diverge. (C) QMEANDisCo score (Level of confidence) in RocC prediction in Figure 1A.

chymotrypsin, which cleaves C-terminal to aromatic or bulky hydrophobic amino acids (Figure 3-2A) (Wilkins et al., 1999). Different ratios of trypsin and chymotrypsin were incubated with high purity of full-length RocC from 1:100 (RocC:protease ration) to 1:5,000 at room temperature for 30 minutes. The digested products were separated by SDS-PAGE (Figure 3-2B). Trypsin digestion resulted in extensive degradation of RocC at the highest concentrations, however, chymotrypsin treatment failed to produce total digestion even in the highest ratio, 1:100. Further digestion A time course of trypsin digestion at the 1:2,500 ration revealed a single trypsin stable species of ~11kDa (Figure 3-2C).

Bioinformatics tools coupled with the limited proteolysis results enabled us to predict the approximate boundaries of the stable FinO domain for generating samples for further binding assays and crystallization. We used Psipred (Buchan et al., 2013) to predict the RocC secondary structure, which suggested the N-terminus of RocC as a highly α-helical structure (10-140). This N-terminal FinO core has a predicted unstructured C-terminus (Figure 3-3). A combination of biochemical and bioinformatic clues allowed us to design 1-137, 14-110, 14-126, 14-137, 24-110, 24-126, 24-137 constructs, each containing the FinO domain and close to 11 kDa in size (Figure 3-4A).

3.3.2. RocC:RocR interaction requires $\mathrm{RocC}_{14\text{-}23}$

We carried out FP assays with 5'-FAM labeled SL3 to define a minimal RNA-binding region of RocC that would be amenable to structural studies. Limited proteolysis combined with bioinformatics suggested $\mathrm{RocC}_{24\text{-}126}$ could be a stable folded core; however, this construct was defective in interacting with SL3 (Figure 3-5A). In contrast, $\mathrm{RocC}_{14\text{-}126}$ displayed a similar binding affinity to $\mathrm{RocC}_{1\text{-}126}$ or $\mathrm{RocC}_{1\text{-}137}$, indicating that residues 14-23 contain critical residues for its interaction. On the other hand, 1-13 or 127-137 appeared to be dispensable for RNA recognition. From this result, $\mathrm{RocC}_{14\text{-}126}$ was used for further EMSA and FP studies.

3.3.3. The upper part of the stem is dispensable for RocC:RocR interaction

Previous work on the FinO:FinP system indicated that FinO recognizes the base of the terminator hairpin and the 3' ssRNA tail (Arthur et al., 2011; Jerome & Frost, 1999). The

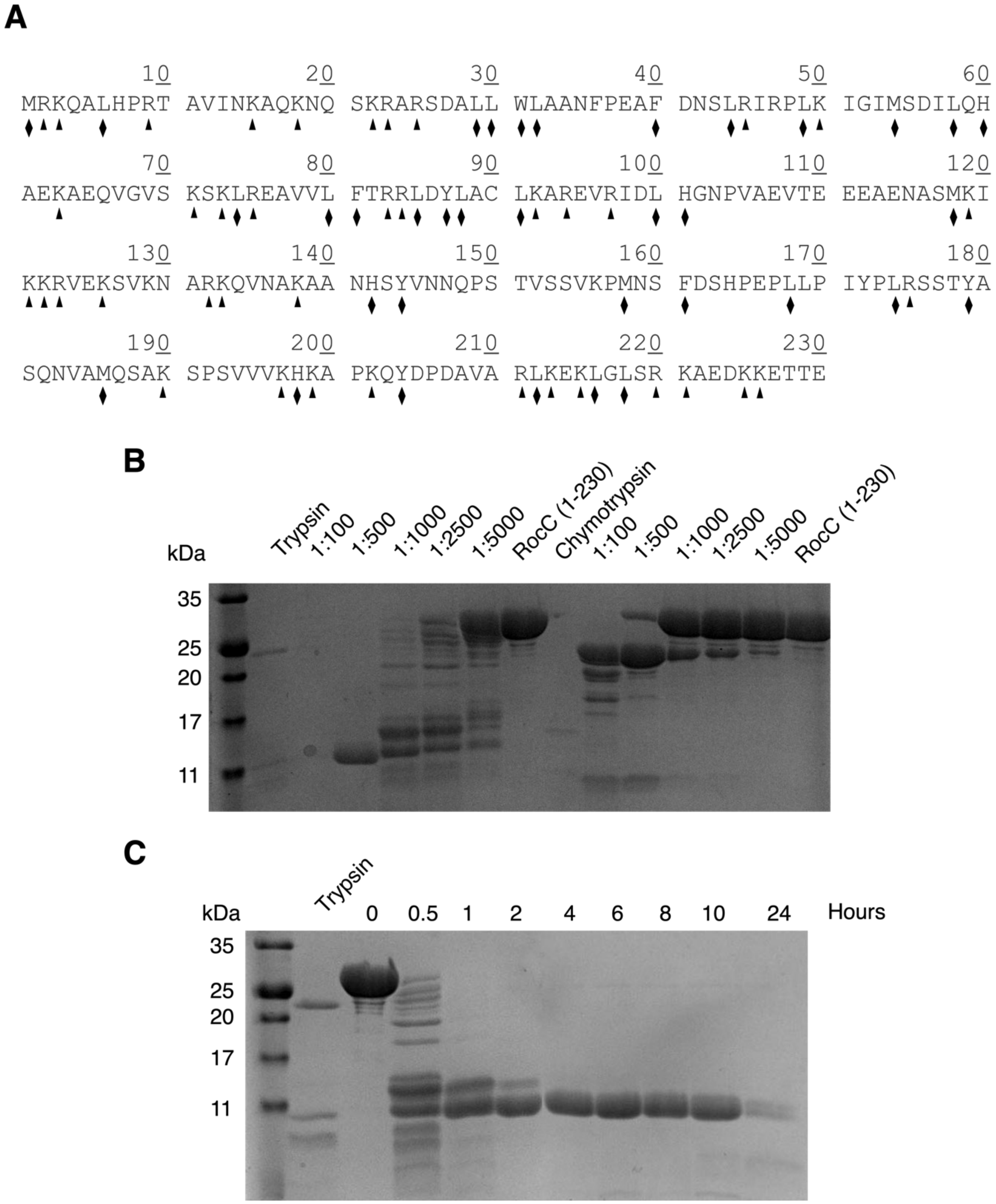
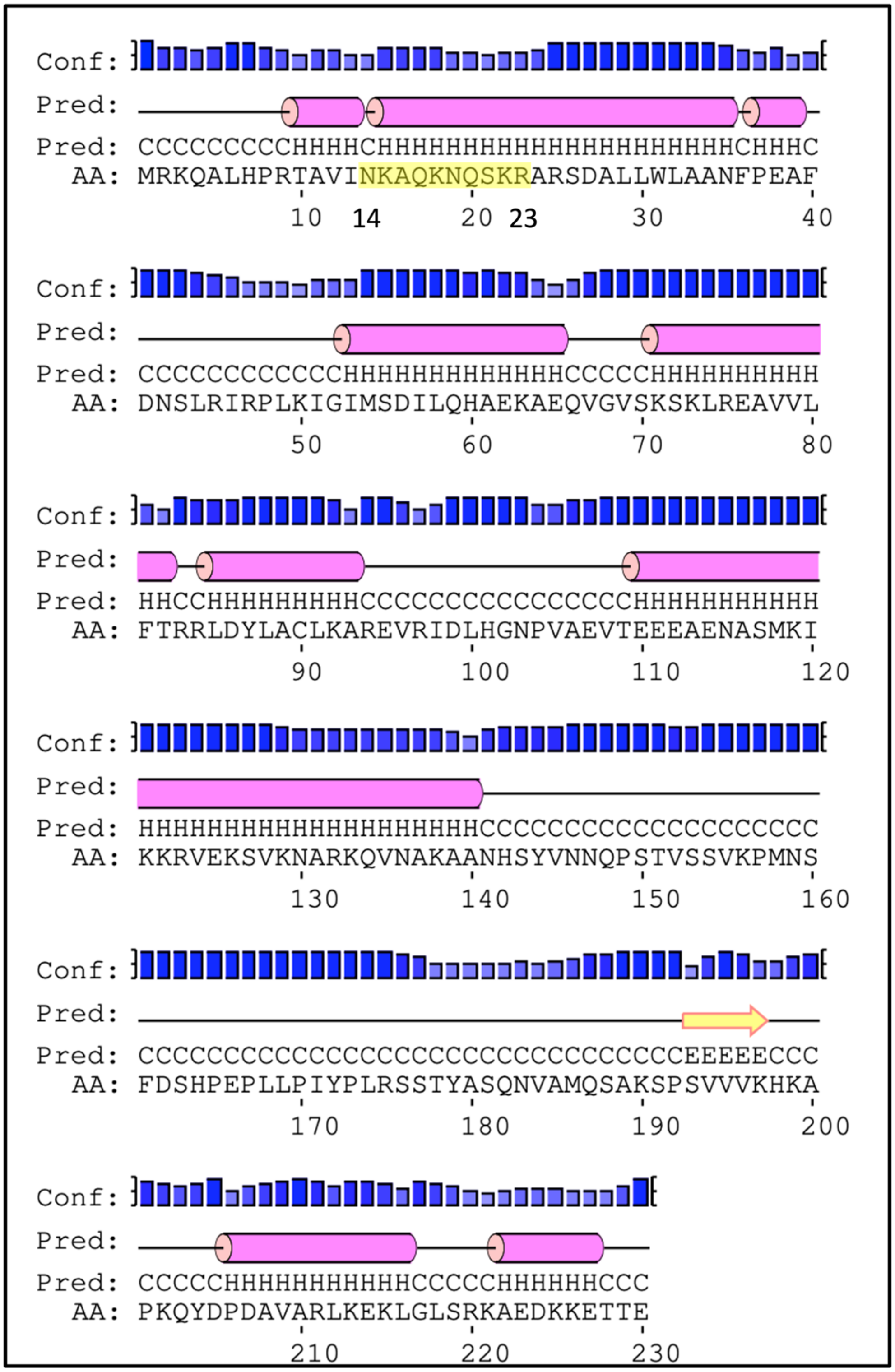


Figure 3-2. Limited proteolysis of RocC.

(A) Predicted digestion sites in RocC. Triangles and diamonds indicate trypsin and chymotrypsin sites, respectively. (B) RocC digestion with trypsin and chymotrypsin in different ratios. (C) RocC digestion time course with a trypsin ratio of 1:2,500 at room temperature.



```
Conf:
Pred:
Pred: CCCCCCCCCHHHHCHHHHHHHHHHHHHHHHHHHHHCHHHC
  AA: MRKQALHPRTAVINKAQKNQSKRARSDALLWLAANFPEAF
              10  14      20  23      30      40

Conf:
Pred:
Pred: CCCCCCCCCCCCCHHHHHHHHHHHHCCCCCHHHHHHHHHH
  AA: DNSLRIRPLKIGIMSDILQHAEKAEQVGVSKSKLREAVVL
              50        60        70        80

Conf:
Pred:
Pred: HHCCHHHHHHHHHCCCCCCCCCCCCCCCHHHHHHHHHHHH
  AA: FTRRLDYLACLKAREVRIDLHGNPVAEVTEEEAENASMKI
              90       100       110       120

Conf:
Pred:
Pred: HHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCC
  AA: KKRVEKSVKNARKQVNAKAANHSYVNNQPSTVSSVKPMNS
             130       140       150       160

Conf:
Pred:
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEEEECCC
  AA: FDSHPEPLLPIYPLRSSTYASQNVAMQSAKSPSVVVKHKA
             170       180       190       200

Conf:
Pred:
Pred: CCCCCHHHHHHHHHHHCCCCCHHHHHHCCC
  AA: PKQYDPDAVARLKEKLGLSRKAEDKKETTE
             210       220       230
```

Figure 3-3. Predicted secondary structure of RocC using PSIPRED.

The blue bar indicates confidence in the prediction. The pink cylinder shows α-helix, yellow indicates is β-sheet. Yellow highlight indicates $\text{RocC}_{14\text{-}23}$ residues.

ProQ/FinO domain of *E. coli* ProQ showed 8 nM binding affinity to a derivative of FinP SLII lacking the loop, corroborating that the FinO/ProQ domain interacts with bottom part of the intrinsic terminator (Chaulk et al., 2011). More recent studies have similarly shown that the FinO domain of *E. coli* ProQ binds similar hairpin-tail structures with at least a 3 base-pair stem and a 4 nucleotide 3' tail, consistent with the notion that ProQ, like FinO, binds mainly the base of the stem and 3' tail (Stein et al., 2020). Similarly, RocC was shown to bind to the rho-independent terminator of RocR, SL3 (Attaiech et al., 2016). SL3 consists of an eleven base-pair stem with a 7 nucleotide loop, and a five nucleotide 3' tail (Figure 3-6A). To test the relative importance of the RocR SL3 loop, stem, and 3' tail, we generated various mutants in these areas and assessed their ability to bind RocC by EMSA (Figure 3-4B, C, D, E). Replacement of the 7 nucleotide SL3 loop, with a UNCG tetraloop sequence did not impact binding affinity, suggesting the loop is not an important determinant for RocC:RocR interactions (Figure 3-6C). We designed a series of constructs to test the importance of stem length, which is truncated gradually by two base pairs from the loop (Figure 3-5A). Deletion of the stem down to 5 base-pairs did not impact binding affinity. However, deletion of the stem down to 3 base pairs resulted in a 3-fold lower binding affinity than WT (Figure 3-4C, 3-5C). Together, we concluded that RocC requires at least 5 base-pair stem to maintain tight binding affinity.

3.3.4. The length of the tail is essential for RocC:RocR binding

Many studies have demonstrated the importance of the 3' tail of Rho-independent terminators for interactions with FinO family proteins (Arthur et al., 2011; Bauriedl et al., 2020; Holmqvist et al., 2018; Jerome & Frost, 1999). The effect of changes in the tail length on ProQ binding demonstrated that a 4 nucleotide tail was a minimal requirement to maintain binding affinity similar to WT (Stein et al., 2020). We generated SL3s with longer or shorter 3' tails and EMSA showed that even deletion or extension of the tail by a single residue resulted in a marked reduction in binding affinity (Figure 3-4D, 3-5D, 3-6B). The dissociation constant of $\mathrm{RocR_{6nt}}$ was too low to be calculated and $\mathrm{RocR_{4nt}}$ showed 5.5 times lower binding affinity than WT, $\mathrm{RocR_{SL3}}$. Even though the K_D value of $\mathrm{RocR_{3nt}}$ and $\mathrm{RocR_{7nt}}$ showed tighter binding than $\mathrm{RocR_{4nt}}$, they were not accurate since their binding fraction was below 40%. Removal of the 3' ssRNA tail altogether completely eliminated RocC:RocR interactions. The importance of the 3' tail length led us to test

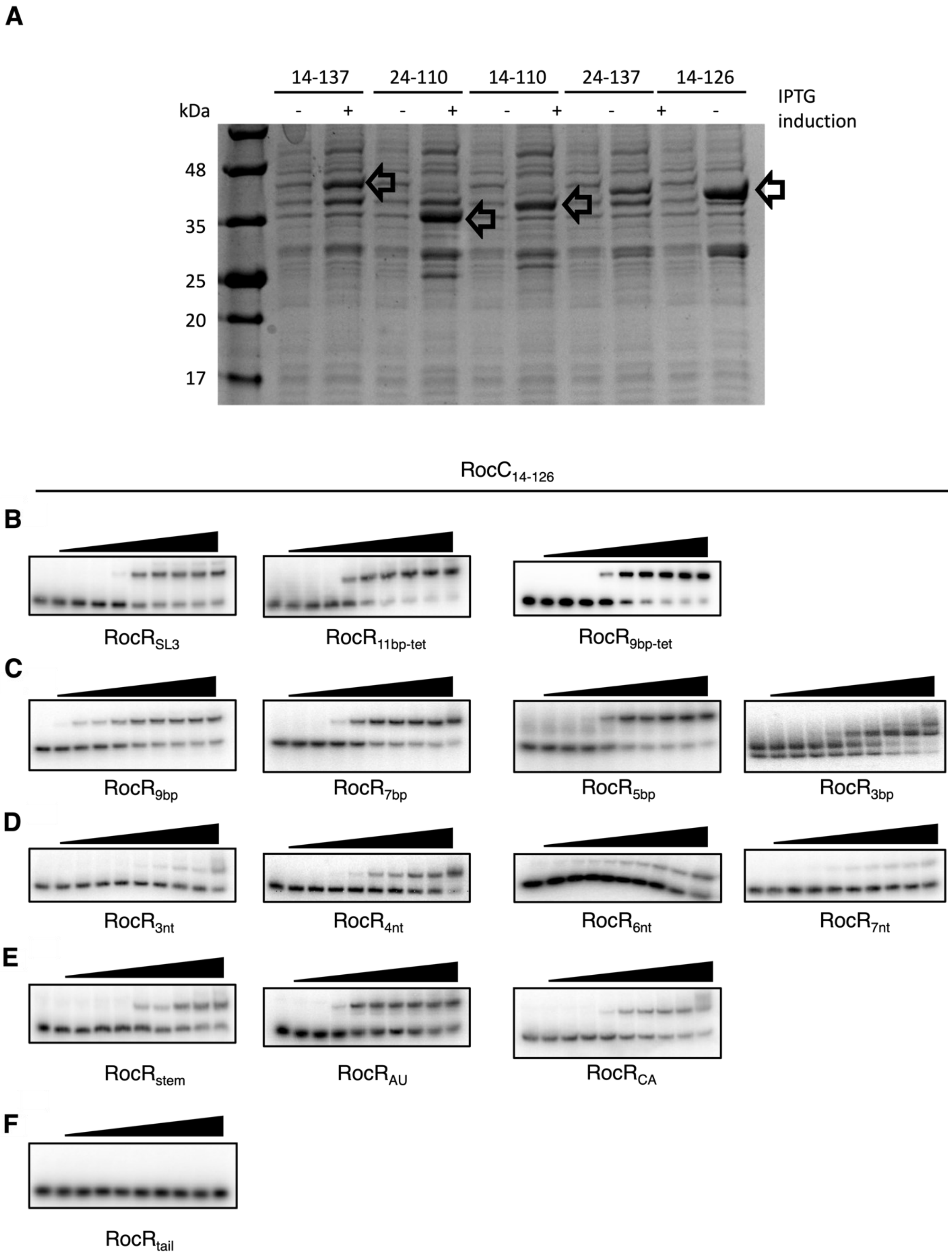

A
14-137
24-110
14-110
24-137
14-126
IPTG induction
kDa
- + - + - + - + -
48
35
25
20
17
RocC$_{14-126}$
B
RocR$_{SL3}$
RocR$_{11bp-tet}$
RocR$_{9bp-tet}$
C
RocR$_{9bp}$
RocR$_{7bp}$
RocR$_{5bp}$
RocR$_{3bp}$
D
RocR$_{3nt}$
RocR$_{4nt}$
RocR$_{6nt}$
RocR$_{7nt}$
E
RocR$_{stem}$
RocR$_{AU}$
RocR$_{CA}$
F
RocR$_{tail}$

Figure 3-4. Expression test of different RocC constructs and results of binding tests using EMSA.

(A) SDS-PAGE gel to test expression level of different RocC deletion mutants. + indicates IPTG induction and – means before induction. Arrows indicates overexpression of designated constructs. (B-F) Representative EMSA gels with the indicated RocC and RocR constructs. The lanes represent 2-fold serial dilutions of RocC concentration gradient from 32 μM (10th lane) to 0.125 μM (2nd lane) (B) Comparison of the effect of loop and stem composition on binding affinity (C) Analysis of stem length on binding affinity. (D) Effect of 3' ssRNA tail length on binding affinity. (E) Effect of substitution mutations on binding affinity. (F) Effect of 5nt tail of RocR on binding affinity.

A

RocC	Mutant	K_D (μM)	Standard error (μM)
1-126	WT	0.216	0.017
1-137	WT	0.30	0.05
14-126	WT	0.34	0.06
24-126	WT	1.10	0.32

B

	K_D (μM)		K_D (μM)
RocR_{SL3}	1.51 ± 0.07	$\text{RocR}_{\text{stem}}$	2.76 ± 0.37
RocR_{3bp}	4.24 ± 0.57	RocR_{AU}	0.66 ± 0.05
RocR_{5bp}	1.04 ± 0.05	RocR_{CA}	3.21 ± 0.18
RocR_{7bp}	0.89 ± 0.06	RocR_{P}	2.20 ± 0.23
RocR_{9bp}	1.27 ± 0.15	$\text{RocR}_{\text{11bp-tet}}$	1.56 ± 0.10
RocR_{3nt}	3.50 ± 3.91	$\text{RocR}_{\text{9bp-tet}}$	1.69 ± 0.08
RocR_{4nt}	8.45 ± 3.29		
RocR_{6nt}	ND		
RocR7_{nt}	4.56 ± 1.45		

Table 3-1. Binding affinities measured using EMSA. (A) Binding affinities for the indicated RocC:RocR_{SL3} complexes determined by FP (see Figure 3-5A). (B) Binding affinities for the indicated RocR_{SL3} bound to $\text{RocC}_{14\text{-}126}$ as determined by EMSA (see Figure 3-5B, C, D, F).

whether the 5 nucleotide tail by itself ($\text{RocR}_{\text{tail}}$) is sufficient to bind RocC, however this RNA did not interact with RocC under EMSA conditions (Figure 3-4F). This work indicated that the FinO domain of RocC preferentially binds an SL3 with a minimum 5 base pair stem and a 5 nucleotide 3' single stranded tail.

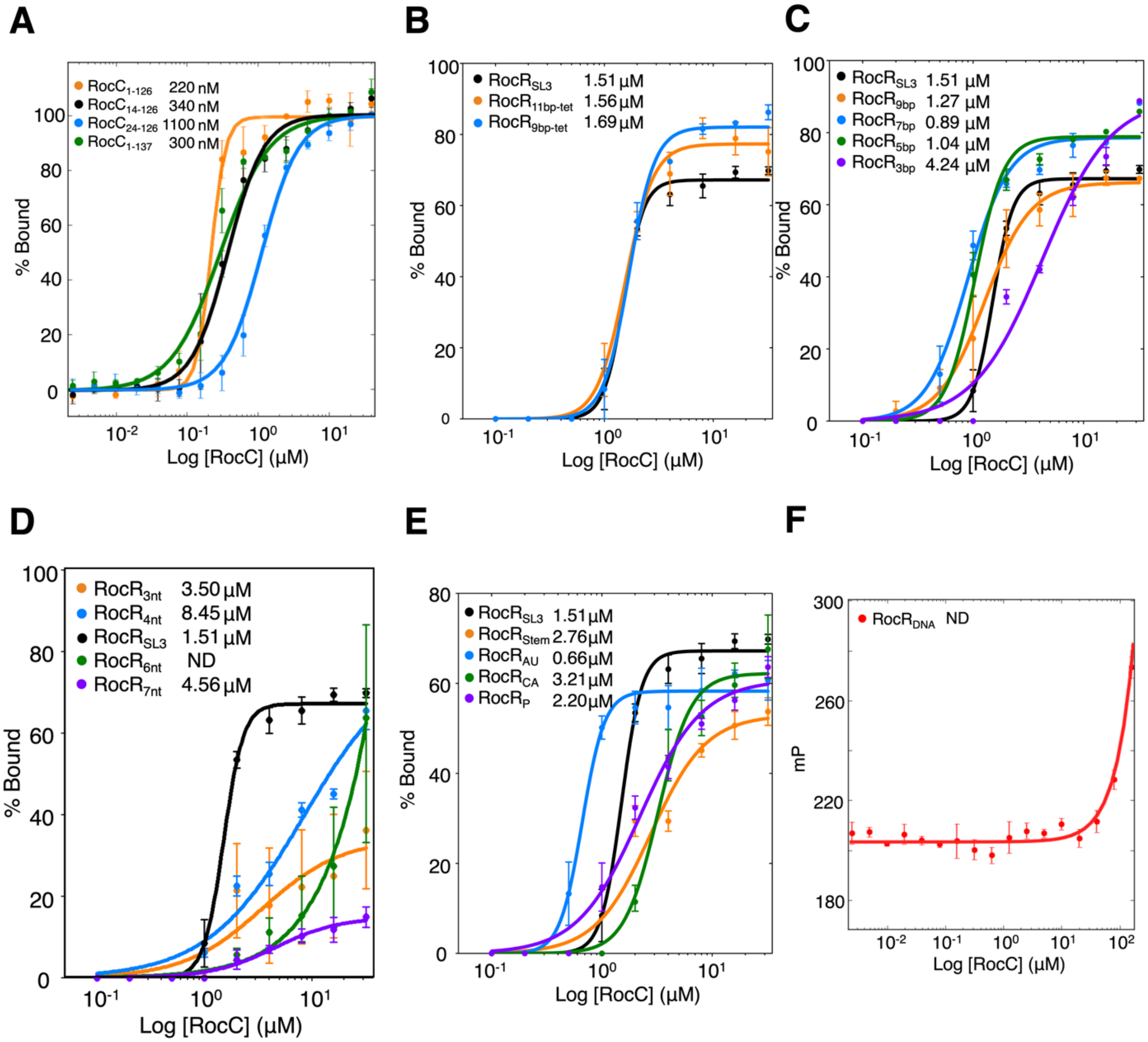


Figure 3-5. Binding analyses of various RNA constructs with different FinO domains of RocC using fluorescence polarization (FP) and electrophoretic mobility shift assay (EMSA).

(A) FP binding assay for 5’ FAM-labeled SL3 with different FinO domains of RocC. (B), (C), and (D) EMSA binding assay for RocC$_{14\text{-}126}$ with 5’ radiolabeled SL3 with different loop size, stem lengths, and tail lengths, respectively. (E) EMSA binding assay for RocC$_{14\text{-}126}$ with 5’ radiolabeled SL3 with various substitution mutants. (F) FP binding assay for 5’ FAM-labeled DNA version of SL3 with RocC$_{14\text{-}126}$.

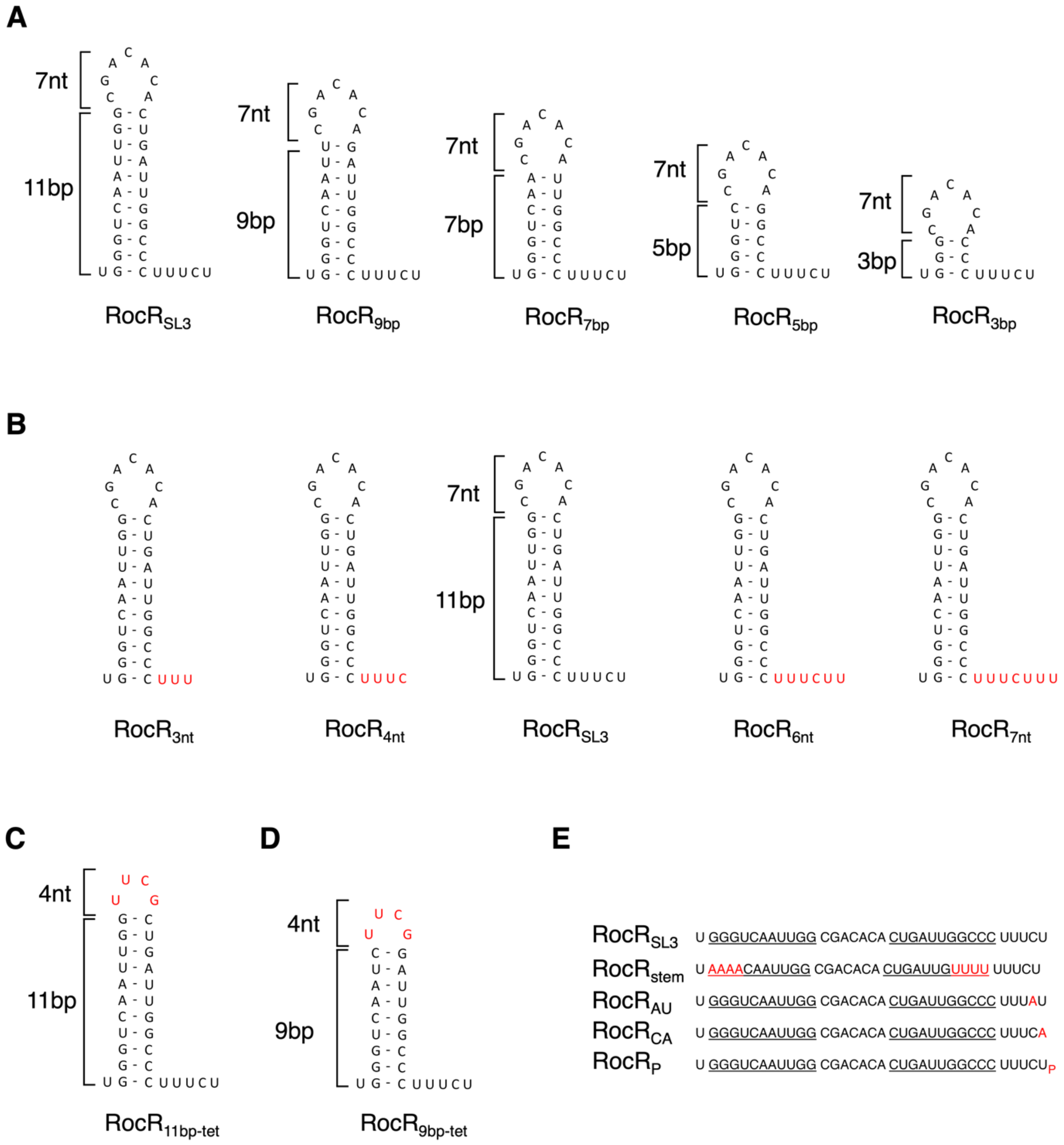


Figure 3-6. Various RNA constructs used for binding tests.

(A) SL3 with different stem lengths. (B) RocR$_{SL3}$ with different tail lengths. (C) RocR$_{SL3}$ with different loop sizes. (D) Predicted folding of RocR$_{9bp\text{-}tet}$, crystallized with RocC$_{14\text{-}126}$ using RNAfold web server. The color gradient indicates confidence in the prediction. (E) Various substitution mutants. Underline shows predicted double-stranded stem. Red indicates substitutions from the original sequence.

3.3.5. RocC:RocR recognition is not sequence-specific

Our EMSAs with different tail lengths demonstrated RocC requires a 5 base-pair stem and especially recognizes the SL3 5nt 3' tail, UUUCU. Previous work suggested that the precise sequence of FinP SLII is not critical for FinO interaction (Jerome & Frost, 1999). Neither substitutions at the base of the stem, nor sequence substitutions in the 3' ssRNA tail resulted in large reductions in binding affinity. Arthur and colleagues tested the importance of the chemical integrity of the 3'-most nucleotide of the tail for interaction of FinP SLII with FinO. Phosphorylation of the 3' nucleotide blocked binding, which could be restored by phosphatase treatment. Binding could be further blocked by $NaIO_4$ oxidation of the terminal ribose, suggesting the importance of the terminal 3' hydroxyl group of FinP in FinO recognition (Arthur et al., 2011).

We were interested to test if RocC bound RocR in a sequence-specific manner. We substituted the GC-rich sequence at the base of the stem with AU base pairs and found that this did not significantly impact the binding affinity. Likewise, the introduction of adenine at either the 4 positions or the terminal 5 positions of the polypyrimidine tail (5'-UUUCU-3') did not significantly impact binding, suggesting a lack of sequence specificity in recognition of this region (Figure 3-4E, 3-5E). In contrast to the results with FinO, however, phosphorylation of the 3' terminal nucleotide of RocR SL3 does not significantly reduce RocC binding affinity (Figure 3-4E). Interestingly, the SL3 sequence in DNA (U replaced with T) failed to bind RocC (Figure 3-4F). In conclusion, RocC recognizes geometry of the A-form RNA helix with a 5 nucleotide 3' tail.

3.3.6. FinO domain of RocC binds to intrinsic terminator of RocR, 1:1

Ghetu and colleagues performed gel shift assays using FinO and GST-FinO to examine the binding ratio between FinO and SLII of FinP, demonstrating the formation of a one to one complex (Ghetu et al., 1999). From this previous stoichiometry study, we hypothesize that the FinO domain of RocC and SL3 of RocR will show the same 1:1 binding. *In vitro* transcribed SL3 was incubated with either RocC or GST-RocC or a mix of both RocC and GST-RocC for 30 minutes. If RocC binds RocR as a 1:1 complex, then we would expect to see complex bands corresponding to RocC:RocR or GST-RocC:RocR. However, if RocC instead bound the RNA as a dimer, we would expect to see three shifted species in the RocC + GST-RocC sample corresponding to the two homodimers, as well as a RocC:GST-RocC heterodimer bound to the RNA. The results clearly show that no heteromeric complex is observed in the sample in which RocC and GST-RocC are mixed, consistent with a 1:1 RocC:RocR binding stoichiometry (Figure 3-7A, B). The oligomeric states of individual components, truncated RocC (non-GST tagged) or RocR, and protein:RNA complexes were confirmed by SEC-MALS (size exclusion chromatography-multi angle light scattering) (Figure 3-7C). Individual components showed single peak through SEC and the measured molecular weight indicated that RocC and RocR bind in a 1:1 stoichiometry in solution (Figure 3-7E).

3.4. Discussion

We showed here that the FinO domain of RocC containing a short, proteolytically sensitive N-terminal region is necessary and sufficient for high affinity interactions with $\mathrm{RocR_{SL3}}$. The deletion of ten residues in a range of 14-23 weakened binding to $\mathrm{RocR_{SL3}}$ ~3-fold, suggesting that these ten residues are important to form a stable complex. The extensive screening with variously modified SL3 constructs concluded that SL3 requires at least a 5 base-pair stem, and a 5 nucleotide 3’ tail to be recognized by RocC. RocC and RocR SL3 were shown to form a stable 1:1 complex using modified EMSA and SEC-MALS.

3.4.1. $\mathrm{RocC_{14\text{-}126}}$ is a minimal protein domain that tightly binds to SL3

Previous deletion studies using FinO stressed the importance of the FinO domain as an RNA binding core (Arthur et al., 2003; Ghetu et al., 1999). Analysis of the binding affinities of several RocC truncation mutants to SL3 using FP provided insight into the regions of RocC that are important for interactions with SL3. We did not observe dramatic differences between $\mathrm{RocC_{1\text{-}126}}$, $_{14\text{-}126}$, and $_{1\text{-}137}$ binding affinities, even though the $\mathrm{RocC_{1\text{-}126}}$ construct used in previous EMSA study (Attaiech et al., 2016), showed slightly higher binding affinity than the others. An interesting result was observed comparing $\mathrm{RocC_{14\text{-}126}}$ and $\mathrm{RocC_{24\text{-}126}}$. While $\mathrm{RocC_{24\text{-}126}}$ corresponded to trypsin-stable folded core, it displayed ~3 times weaker binding affinity than the larger construct, indicating a role for the apparently flexible 14-23 region. $\mathrm{RocC_{24\text{-}126}}$ was a highly soluble construct (maximum solubility was 110 mg/mL) and stable over the purification, suggesting the defect in binding was not a folding problem, but instead was the direct lack of the 14-23 region. Previous studies showed that deletion of the extended N-terminus in FinO$_{1-45}$ actually enhances FinO binding to SLII of FinP, however this deletion resulted in a functional loss *in vivo* (Arthur et al., 2003). Four out of ten residues in 14-23 are positively charged residues, consistent with the idea that this region might make electrostatic interactions with RNA (Figure 3-3).

3.4.2. RocC requires a minimal 5 base-pair stem base and 5 nucleotide 3' tail for optimal binding

Our RNA mutagenesis analysis indicates that the 5 base pairs at the base of the SL3 hairpin, as well as the 5 nucleotide single stranded 3' tail are critical elements for recognition. Extending the length of the stem, or modifying the sequence of the loop to the stable UNCG tetraloop did not significantly affect binding affinity. Our finding that the thermodynamically stable (Bottaro & Lindorff-Larsen, 2017) UNCG tetraloop was compatible with binding was of interest as this relatively rigid structure might be an aid for our future crystallization efforts. Unlike coaxial stacking of A form helices, a flexible loop could interrupt crystal packing so that replacing the exposed loop with stable tetraloop is one of strategies to increase the chance of protein:RNA complex crystallization (Golden, 2007; Reyes et al., 2009). Our results showed that replacement of the loop, and even removal of two base-pairs from stems closer to the loop, does not influence their interactions with RocC (Figure 3-4B, 3-5B). This result was consistent with previous findings in the FinOP system. Sequence alignment of FinPs from various plasmids showed the double stranded stem to be highly conserved, however the unpaired loop is not conserved, and it has been suggested that this result suggests that the unpaired loops are not contacted by FinO and instead are critical for FinP-mRNA binding (Frost et al., 1994).

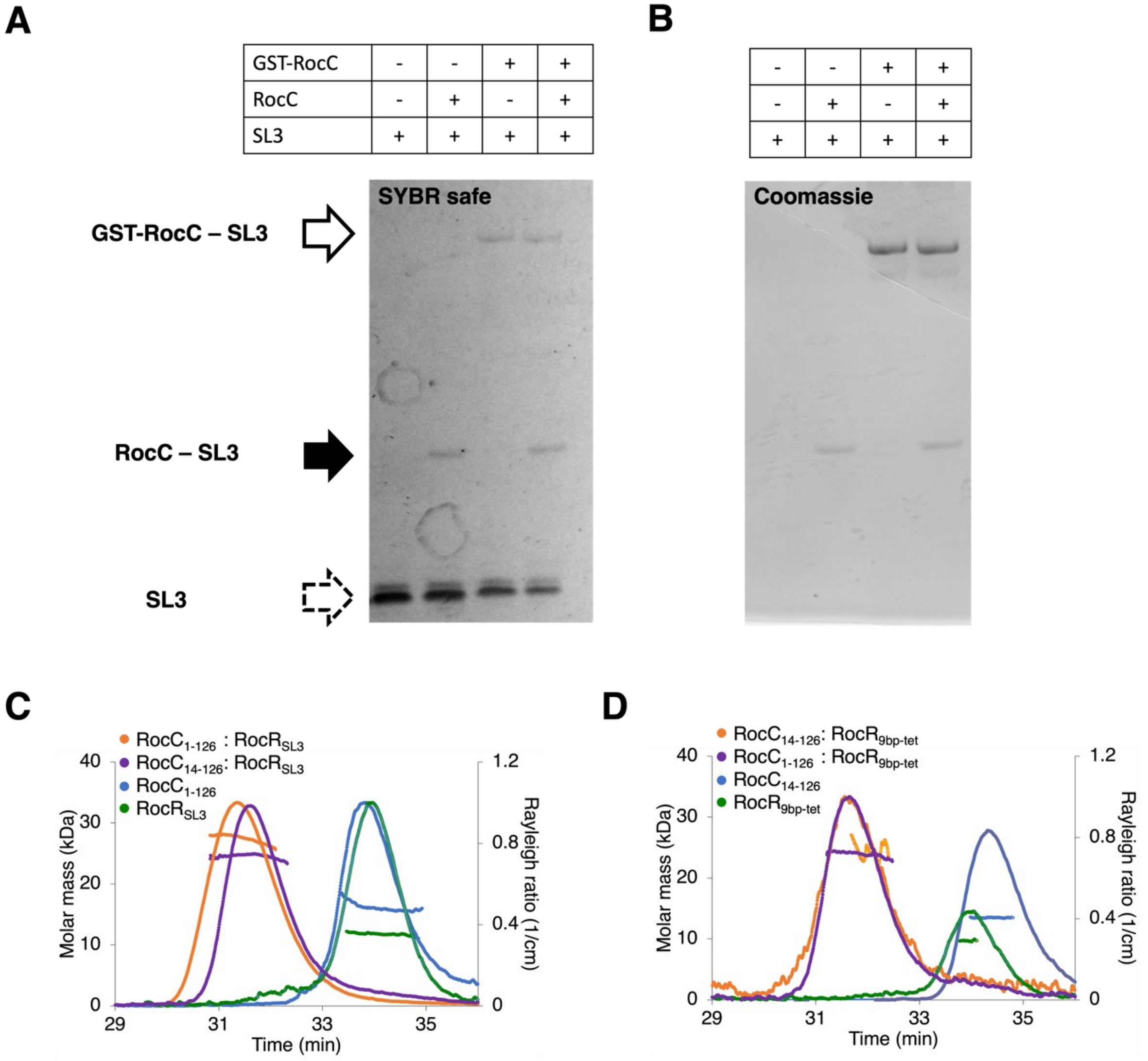
A striking finding from our present work is that RocC selectively binds to Rho-independent terminator structures with a 5 nucleotide 3' polypyrimidine tail (Figure 3-4D, 3-5D). Reduction or elongation of the tail length by just one nucleotide resulted in dramatic deterioration in binding, with many of the mutated RNAs failing to demonstrate binding saturation under the EMSA conditions. The effect of tail length on the binding of *E. coli* ProQ for one of its natural targets, the *malM* 3' UTR revealed somewhat different conclusions (Stein et al., 2020). In this case, the natural *malM* 3' UTR has a 7 nucleotide 3' tail, however changes to the tail length from 4 to 9 nucleotides did not affect binding affinity, but, deletion of the tail to 3 nucleotides or less led to a loss in ProQ binding (Stein et al., 2020). Hfq is another RNA chaperone in bacteria which is known for recognizing the intrinsic terminator shape of sRNAs to regulate gene regulation. Hfq represses expression of the *ptsG* gene through the SgrS sRNA in response to glucose-phosphate stress in *E. coli*. One study showed that by shortening the 3' tail from 8 nucleotides (WT) to 4 nucleotides, the repression of *ptsG* mRNA was gradually reduced due to weaker binding (Otaka et al., 2011b).

Co-immunoprecipitation coupled with RNA deep sequencing revealed RocR as the primary RocC binding partner in *L. pneumophila* (Attaiech et al., 2016). The fact of this high selectivity presented the possibility that RocC might bind its RNA target in a sequence-specific manner. The RocR 3' tail is not pure penta-uridine, but has one cytosine at the penultimate position (5'-UUUCU-3'), which we reasoned might provide binding specificity. Mutation of this nucleotide to adenosine resulted in an ~1.5-fold reduction binding affinity, indicating that a cytosine (or even a pyrimidine) might be selected for at this position. We also tested the importance of the sequence in the lower stem. Mutation of the 4 G-C pairs in the region to A-U weakened binding ~2-fold compared to $\mathrm{RocR_{SL3}}$, suggesting at best only a minor role for sequence specificity. Our previous work on FinO (Arthur et al., 2011) indicated that the integrity of the 3' terminal nucleotide in FinP is especially critical for FinO binding. In RocR, the terminal nucleotide is a U and mutation of this residue to an A only reduced binding affinity ~2-fold. Contrary to our previous results in the FinO system, phosphorylation of the terminal residue only showed a minor, ~1.5-fold reduction in binding affinity for RocC:RocR binding. However, we note that this experiment was performed with T4 PNK-labeled RNA. T4 PNK is also a 3' phosphatase, and therefore it is very possible that the 3'-phosphate might have been removed in that experiment. We did not see any RocC binding to a DNA version of RocR SL3, indicating that RocC must sense the ribose sugars of the RNA, either directly, or potentially indirectly through geometric differences between the RNA A-form helix and the DNA B-form helical structure (Figure 3-5F).

3.4.3. FinO domain of RocC and SL3 of RocR form a 1:1 complex

Our stoichiometry study showed that the FinO domain and its target sRNA binds in a 1:1 complex, both via an EMSA experiment as well as by SEC-MALS (Figure 3-7). Similar experiments were previously performed with full-length FinO and its target region SLII of FinP (Ghetu et al., 1999). A recent study of *E. coli* ProQ, which has 34% sequence similarity with RocC, reported that not only the FinO domain, but also the ProQ Tudor domain could contact RNA, opening up the possibility for more complex binding modes (Gonzalez et al., 2017). Evidence for this came from HDX experiments which showed that parts of the ProQ linker and Tudor domains are protected by interacting with SraB, one of the ProQ RNA targets, indicating that linker and C-terminal domain are involved in RNA binding event. However, it is still unknown if either RocC or ProQ lacking the FinO domain can act as an independent RNA binding protein.

Together, our work identified minimal RocC and RocR constructs that form a stable complex. This established the foundation of sample preparation for our further structural and biochemical analysis of this system.



E

	Measured M.W. (kDa)	Theoretical M.W. (Da)	M_w/M_n	Oligomeric state
$\text{RocC}_{1\text{-}126}$	$1.633\text{x } 10^4$ (± 0.036%)	15938.5	1.002 (± 0.023%)	Monomer
$\text{RocR}_{14\text{-}126}$	1.349×10^4 (±0.019%)	13252.32	1.000 (±0.028%)	Monomer
RocR_{SL3}	$1.180 \text{ x } 10^4$ (± 0.013%)	11355.6	1.000 (± 0.019%)	Monomer
$\text{RocR}_{\text{9bp-tet}}$	9.661×10^3 (±7.121%)	9067.2	1.000 (±10.077%)	Monomer
$\text{RocC}_{1\text{-}126}$: RocR_{SL3}	$2.743 \text{ x } 10^4$ (± 0.790%)	27294.1	1.001 (± 1.120%)	Heterodimer
$\text{RocR}_{1\text{-}126}$: $\text{RocR}_{\text{9bp-tet}}$	2.411×10^4 (±0.020%)	25005.72	1.000 (±0.028%)	Heterodimer
$\text{RocR}_{14\text{-}126}$: RocR_{SL3}	2.463×10^4 (±0.020%)	24607.92	1.000 (±0.028%)	Heterodimer
$\text{RocR}_{14\text{-}126}$: $\text{RocR}_{\text{9bp-tet}}$	2.422×10^4 (±18.863%)	22319.52	1.002 (±26.606%)	Heterodimer

Figure 3-7. Stoichiometry study between FinO domain of RocC and intrinsic terminators.

^{32}P-labeled SL3 of RocR was incubated with either RocC$_{1\text{-}126}$ or GST-RocC$_{1\text{-}126}$ in ice for 30 minutes. Complex formation was tested by native gel electrophoresis and the bands corresponding to RocR$_{SL3}$, RocC$_{1\text{-}126}$:RocR$_{SL3}$, and GST-RocC$_{1\text{-}126}$:RocR$_{SL3}$ are indicated. (A) Native gel stained with SYBR safe. (B) Native gel stained with Coomassie blue. (C) and (D) SEC-MALS results demonstrating that RocC forms a 1:1 complex with RocR$_{SL3}$. (E) Predicted and experimental masses as determined by SEC-MALS.

Chapter 4.

Structural understanding of how RocC recognizes the RocR intrinsic terminator

4.1. Overview

The FinO family is an RNA binding protein in proteobacteria that recognizes its target RNAs through its FinO domain. Despite the emerging importance of the FinO family as a post-transcriptional regulator associated with sRNAs, there is a lack of understanding in how members of the FinO family recognize their binding partners. To understand the mode of binding of the FinO domain:RNA complex, we tried to crystallize one of the FinO family, RocC, with its specific binder, RocR. Our systematic approaches, including buffer screening and additive screens, produced diffracting crystals which we used to determine the structure of two apo-RocC and one RocC:RocR complex using X-ray crystallography. The complex structure reveals specific recognition of the 3' nucleotide of the terminator by a conserved pocket involving a β-turn-α-helix motif, while the hairpin portion of the terminator is recognized by a conserved α-helical N-cap motif. The structure of the complex was also investigated using SEC-SAXS. The excellent agreement of the crystallized complex with the SEC-SAXS model suggests that the FinO domain in complex with the intrinsic terminator is static in solution. Taken together, these data reveal the molecular mechanism of how the FinO domain recognizes its binding partner RNA.

4.2. Introduction

The FinO family of proteins is defined by the FinO (PF04352) domain. The FinO domain is a small, α-helical structure of around 100 amino acids with a positively charged, concave side that is thought to be a primary RNA interaction surface (Ghetu et al., 2000; Gonzalez et al., 2017; Immer et al., 2020). Multiple FinO domain proteins have now been shown to bind to Rho-independent transcription terminators, and in the case of FinO and ProQ, these interactions have been shown to be mediated through this conserved concave surface (Arthur et al., 2011; Ghetu et al., 2002). Phylogenetic analysis showed that FinO domain-containing proteins are widely distributed across proteobacteria and FinO domains are usually flanked by flexible N- and/or C-terminal extensions (Attaiech et al., 2016, 2017)'(Attaiech & Charpentier, 2016). Exceptions include NMB1681 in *N. meningitidis* (Chaulk et al., 2010) and Lpp1668 in *L. pneumophila* (Immer et al., 2020) which only consist of the FinO domain. Pioneering work on FinO indicated that the folded FinO domain is responsible for RNA binding, while its N-terminal flexible extension is involved in chaperone activity (Arthur et al., 2003). To date, there have only been structures determined of the isolated FinO domains of FinO, NMB1681 and ProQ (Chaulk et al., 2010; Ghetu et al., 2000; Gonzalez et al., 2017).

A question that has not been answered is the mode of sRNA recognition by the FinO family to conduct sRNA-mediated gene regulation. The determination of the FinO protein structure (Ghetu et al., 2000) enabled the design of FinO mutants to test what surfaces of the protein were in contact with the target SLII RNA through site-specific crosslinking experiments (Ghetu et al., 2002). In addition, a gel-based FRET technique (gelFRET) was used to assess the relative proximity of different residues on the surface of FinO from either the base or the top of the stem of the SLII RNA. These studies showed that most residues that bind to SLII are distributed on positively charged patches and that these regions are in closest contact with the base of the stem and 3’ single-stranded tail (Ghetu et al., 2002).

RNase footprinting of the SLII and SLIIc in the presence and absence of the FinO protein demonstrated that FinO protects the base of the stem and the 3’ single-stranded tail (Arthur et al., 2011). Interestingly, the level of protection of the stem base on the 5' and 3' sides was different. SLIIc does not show any protection on the 5' side, and SLII showed some protection on the 5' side, but it was only a limited area compared to the 3' side. In both RNA samples, the 3' tail was protected. RNase footprinting concluded that the FinO protein interacts with its target RNAs through the 3' tail and lower stem, especially the 3' side of the stem. A FinO:SLII model was suggested using SAXS and restrained based on the biochemical studies. The model suggested the 3' tail was bound to the positively charged pocket, and the stem was located along the elongated N-terminal α-helix of FinO.

An integrated structural approach including SAXS was also used to probe the mode of binding of *E. coli* ProQ with one of its target RNAs, SraB (Gonzalez et al., 2017). The ProQ contains a different domain structure from FinO protein. Unlike FinO, ProQ has the FinO domain in the N-terminus, connected via a flexible linker to a Tudor-like domain within the C-terminus. The Tudor domain adopts an SH3-like fold and, in eukaryotes, interacts with methylated lysine and arginine. Hydrogen deuterium exchange (HDX) protection suggested that not only the FinO domain, but also the Tudor domain may contact RNA.

A bacterial three-hybrid (B3H) approach was further used to test the impact of surface-exposed point mutations on interaction of ProQ with either of two target stem-loop terminator RNAs from SibB and cspE transcripts (Pandey et al., 2020). These results suggested that the two target terminators are bound similarly, however with some potential differences. This suggested a model for interaction in which the conserved concave positive surface was a key interaction region. however, it was also suggested the RNA would wrap around the protein to interact with the opposite surface of the domain.

Thus, over the last two decades there have been several efforts to understand the molecular mechanism of RNA target recognition of the FinO domain. However, the mode of binding at the molecular level is still unclear. The lack of clarity results from a dearth of structural information for the FinO family in complex with target RNAs. As we observed from the history of the structural study in FinO with its partner RNA, it is challenging to obtain structural information of the protein:RNA complex. According to protein data bank (https://www.rcsb.org) statistics on 2022 Feb 04, 186,670 structures were deposited. The structures containing protein and RNA together were 2.23% (4,159), which was lower than the number of the structures containing protein and DNA together (3.52%, 6,572). Among the 4,159 protein:RNA complexes, 1,571, 37.8%, were determined by electron microscopy. Electron microscopy has shown a remarkable development in resolution and the available size of molecules that can be studied. However, current electron microscopy technologies require that determined crystal or NMR structures to be docked into EM density due to the limitation of low resolution. Given the current limitation of EM to determine the atomic structure, molecular mechanism studies are mainly implemented by X-ray crystallography and NMR. Excluding the protein:RNA structures determined by EM leaves only 2,588 structures containing protein:RNA complexes in the PDB. This is only 1.4% of the deposited structures and this reflects how challenging the structural study of protein:RNA complexes can be. Despite these difficulties, we attempted to obtain crystals of the protein:RNA complex by systematic approaches as described below.

4.2.1. The strategy to grow crystals and improve the quality of the crystals

A crystal is an array of regularly repeated and homogenous molecules. In X-ray crystallography, the homogenous molecules are biomacromolecules, such as protein, DNA, or

RNA. Structure determination using X-ray crystallography requires a crystal from the biomacromolecule of interest, which enhances X-ray scattering. Integration of the diffraction intensities, together with the relative phase differences between the different diffraction spots are used in a mathematical process called a Fourier transform to calculate the three-dimensional electron cloud of the biomacromolecule in the crystal. The structure is built and fitted into the electron density, and repetitive refinement finalizes the three-dimensional structure.

The preparation of the homogenous molecule is the start of the crystallization process. There are a number of crystallization methods, and vapour diffusion is the most commonly used technique. The reservoir solution (crystallization condition) and a mixture of the molecule to be crystallized and reservoir solution, are placed in a sealed system. The reservoir solution is a mixture of chemicals comprising a buffer system, salt, and precipitant. The precipitant in the reservoir solution plays a crucial role in decreasing the solubility of the biomacromolecule. The decrease of solubility induces crystal formation. Higher precipitation concentrations of reservoirs promote the evaporation of water in the drop. Evaporation of water results in the incremental increase of protein and precipitant concentration in the drop (Figure 4-1). According to Chayen and Saridakis, crystals develop in two phases: first, the nucleation phase and second, the growth phase (Chayen & Saridakis, 2008). As the phase diagram depicts (Figure 4-1), the concentrations of protein and precipitant in the drop should be stalled in the nucleation zone so that the biomacromolecule can form crystal nuclei. The formation of nuclei will drop the protein concentration, and nuclei move into the growth phase in the metastable zone. The search for the nucleation and metastable zones is challenging, and it requires trial and error by controlling variables. There are innumerable factors that could be tested, and McPherson and Garvira categorized these variables into three categories: physical, chemical, and biochemical variables (McPherson & Gavira, 2014). In the following section, we will discuss the methods to test variables for successful crystallization. In addition, some techniques will be introduced to improve the probability of crystallization and the quality of the crystal.

4.2.2. Generation of the crystallizable constructs

The characterization of the biomacromolecule to be studied should be carried out prior to crystallization. Understanding the molecule and the goal of the structural study is essential to

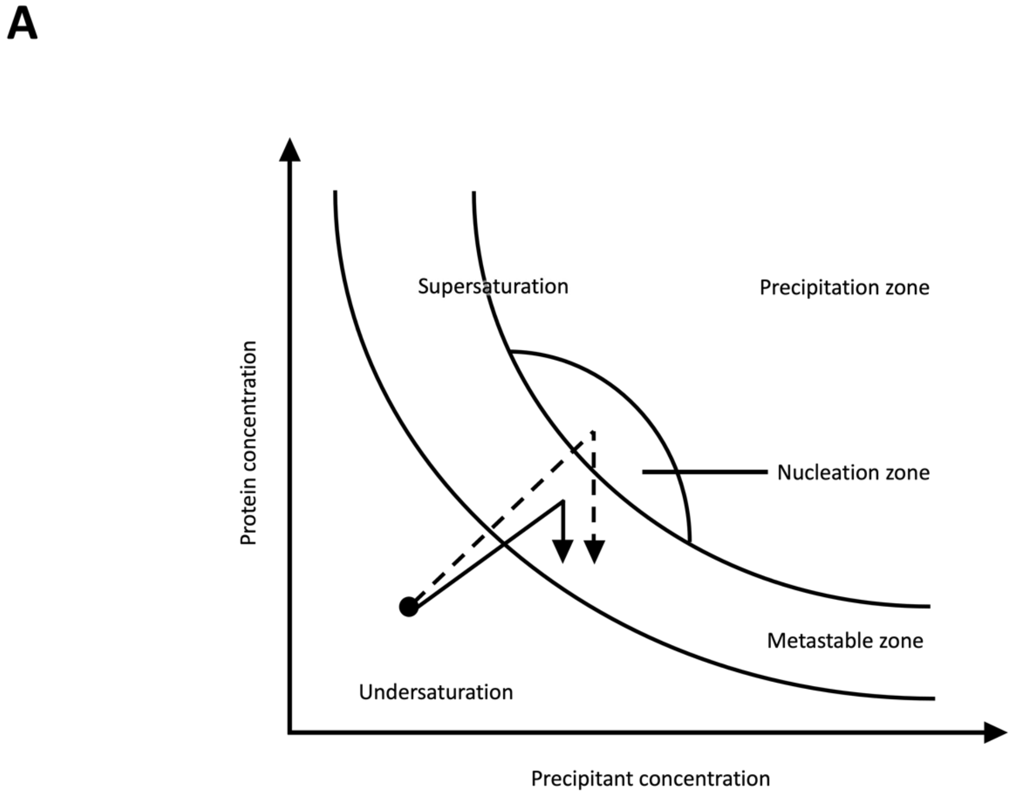


Figure 4-1. Crystallization phase diagram by vapour diffusion.

The black arc indicates the concentration of protein and precipitant in the drop when the crystal plate is set up. Dotted line arrow indicates the protein and precipitant concentration changes in regular setup as time passes. The solid line arrow indicates that the protein and precipitant concentration changes after seeding as time passes.

design the constructs to be crystallized. For example, the inhibitor of apoptosis-stimulating of p53 (iASPP) is a regulatory subunit that confers specificity to the protein phosphatase 1 catalytic subunit (PP-1c). iASPP is comprised of 828 amino acids and secondary structure predictions indicated that the N-terminal ~ 600 amino acids were not structured, and only the C-terminal ~230 amino acids were structured. This C-terminal domain had also been shown to be necessary and sufficient to bind PP-1c. The goal of Zhou and colleagues was to understand the mode of binding of the iASPP:PP-1c complex, and therefore they focused on the structured C-terminal 230 amino acids for structural study. iASPP$_{628\text{-}828}$ in complex with PP-1c was crystallized and the structure was determined (Zhou et al., 2019). The structural study of iASPP:PP1c showed that it is essential to understand the characteristics of biomacromolecules to be crystallized and to generate crystallizable constructs.

The goal of engineering crystallizable constructs is finding the minimal and functional constructs which will answer one's scientific question. These constructs can be designed by truncating dispensable flexible regions, which tends to degrade and disrupt crystallization. Structured protein constructs can be predicted using bioinformatics tools, such as secondary structure prediction or sequence alignment, to reveal conserved regions that might be functionally and structurally important. Limited proteolysis coupled with mass spectrometry is a common method to reveal the boundaries of a folded protein construct. Minimal RNAs for protein:RNA complex formation can be determined using RNase footprinting. Comparison of protected regions on RNA in the absence or presence of the binding partner will distinguish the critical part from the dispensable region for the interaction (Arthur et al., 2011). These minimal RNA constructs also can be explored by studying the binding affinity of truncation mutants (Chapter 3).

4.2.3. Buffer screening to enhance the stability of the biomacromolecules

The buffer solution mimics the physiological condition of the protein in the cell. Most of the experiments are performed *in vitro* after purification, and the choice of buffer solution affects the biomacromolecule of interest in terms of its stability, solubility, activity, etc. The solubility and stability of the protein correlate to the likelihood of yielding crystals. For this reason, the buffer solution should be carefully screened to increase the probability of crystallization. The buffer solution often comprises three components: buffering systems, salts, and additives. The most common choice of the buffer usually has a pKa around pH 7.4, which is the physiological pH. Several buffers, such as Tris, HEPES, phosphate buffers can buffer near pH 7.4. However the influence of the buffer system on the protein's stability is unpredictable, and experiments should be carried out to choose the buffer system, and in many instances proteins have been found to be most stable and crystallize best at pHs significantly different than physiological, thus pH is often a primary variable in the hunt for crystallization conditions. The salt in the buffer solution plays a role in the solubility of the protein. NaCl or KCl are the most commonly used salts, and 150 mM concentration is often chosen to imitate physiological conditions. Low concentration of salt increases protein solubility by interacting with hydrophilic part of protein that helps hydration shell, increasing protein solubility (salting in) (McPherson & Gavira, 2014). Once protein solubility reaches the maximum point enhanced by salt concentration, high salt starts to compete with protein to interact with solvent in an aqueous solution. This competition results in the precipitation of protein molecules, called "salting out". Due to the different natures of proteins, the optimal salt concentration for individual protein stabilization can be quite different. Additives are optional elements depending on the nature of the biomacromolecule. For example, polyamines are commonly used in protein:RNA crystallization to decrease the electrostatic repulsion from the RNA backbone phosphate (Sauter et al., 1999). The choice of buffer solution influences the quality of protein and the success of crystallization, so it should be carefully screened.

The buffer solution is expected to influence protein homogeneity, stability, and solubility, and the characterization of buffer components increases the probability of growing diffracting crystals. Thermal shift assay (TSA) and dynamic light scattering (DLS) are established techniques to find optimal buffer conditions for crystallization from a different perspective. The TSA is a method to screen the thermal stability of the protein in different solution conditions (Ericsson et al., 2006). The sample is mixed with Sypro Orange in the different buffer solutions. Gradual increment of temperature induces protein denaturation, exposing the hydrophobic core. The hydrophobic surfaces buried in a soluble state interact with Sypro Orange. Nonspecific interactions with hydrophobic residues leads Sypro Orange to emit a fluorescent signal. The increasing signal is quenched by the aggregation of proteins which blocks signal emission. The temperature at the midpoint between the highest and lowest temperatures is defined as the melting temperature (T_m). The calculated T_m from the fluorescence signal reflects the thermal stability of the protein depending on different conditions. Comparison of the T_m can identify conditions to maximize the thermal stability of the protein of interest.

Borgstahl described how to use DLS to improve the chance of crystallization and quality of crystal in her methodology publication (Borgstahl, 2007). DLS measures the size of the molecules in the given buffer solutions. The molecule of interest in solution experiences Brownian motion. The collision of the irradiated laser light with molecules scatters the light. Due to the Brownian motion of molecules, the detector records the fluctuating intensity signal and measures it as photons per unit time. The diffusion rate of the molecules depends on their size. The bigger the molecules, the slower the diffusion, which causes less fluctuation signal for bigger particles.

From this data, we can derive the translational diffusion coefficient (D_T), and the hydrodynamic radius (R_H) can be calculated from D_T. In addition, DLS can reveal the polydispersity of the molecule, which allows estimating the probability of crystallization since more monodisperse molecules tend to crystallize more readily than polydisperse samples. Light scattering intensity is proportional to the square of the size of the particle, and thus DLS can be a particularly useful method to determine if the preparation contains aggregated components.

4.2.4. Strategies for obtaining protein in complex with ligand

There are two approaches to obtain protein in complex with ligand: crystal soaking and co-crystallization. The crystal soaking method requires pre-grown crystals. The pre-grown crystal is incubated with the dissolved ligand of interest. In general, the solvent content of protein crystals generally ranges from 27% to 65% (Chruszcz et al., 2008). The presence of solvent channels through most crystals will allow traversing the small ligand through the solvent channel and reaching the active site of the crystallized protein. The soaking method is particularly useful for small inhibitors which can more easily diffuse through crystal solvent channels. The soaking method requires wide solvent channels that the small molecules can travel through. In addition, it is vital to confirm that the binding site is accessible in the crystal lattice. Steric hindrance will prevent the ligand from accessing the binding pocket. Due to the feature of utilizing a pre-formed crystal, structural changes induced by ligand binding will not be detected or could lead to breakage of the crystal.

Co-crystallization is performed by incubating the apo-protein in a solution with a ligand. The formation of the protein /ligand complex prior to crystallization allows the complex to undergo conformational changes induced by the interaction. Theoretically, there is no limit on ligand size because the protein/ligand forms a complex in advance of the formation of a crystal.

4.2.5. Strategy for improving the quality of the crystal

Extensive sparse matrix screening increases the likelihood of obtaining initial crystals. Once a promising condition is found from a matrix screen, further optimization is usually required to increase crystal size and diffraction quality.

Tiny crystals formed from initial screens often occur as “showers” of many crystals, indicating the formation of many nucleation sites. The number of nucleation sites could be decreased by reducing protein concentration. Additives such as 1-5% of glycerol in the drop can be used as a nucleation inhibitor. "Seeding" could also be tried in this case. Seeding is a technique larger crystals can be grown (seeded) from preliminary small crystals. Seeding provides the nucleation sites for crystal growth, often avoiding the crystal shower problem (Figure 4-1). Seeding can be performed in two ways; streaking and microseeding. Crystals should be small to act as seeds so that crystals should be physically crushed. In the streaking method, a cat whisker is often used. The cat whisker is dipped into crushed crystal solution, and the crushed crystals are transferred by streaking on the drop, which is prepared for vapour diffusion. In the microseeding method, the crushed crystals are serially diluted with reservoir solution, and diluted crystals are transferred to the drop for vapour diffusion using a pipette. If the initial screen instead only gives a few very small crystals, this is usually an indication that the protein concentration is too low.

The poorly packed crystal causes an unstable crystal or low-resolution diffraction. There are many aspects to be considered to improve low quality crystals. However, temperature and additive screening are two simple parameters that can be tested. Temperature influences biochemical properties of the protein, such as solubility and stability, and biophysical kinetics, such as the evaporation rate in the crystallization vapour diffusion process. The impact of protein solubility was tested with a set of 30 randomly chosen proteins (Christopher et al., 1998). 86% of these proteins showed temperature-dependent solubility, suggesting temperature is one of the parameters that governs protein solubility. Temperature also affects the water evaporation rate from the droplet in vapour diffusion (Martins et al., 2008). The evaporation rate is associated with the speed of the protein and precipitant concentration changes. Fast evaporation might prevent the formation of crystal nucleation sites in the precipitation zone (Figure 4-1). Slow evaporation could fail to reach the nucleation zone where crystals cannot grow. As described above, additive screens are another option to improve the quality of the crystals.

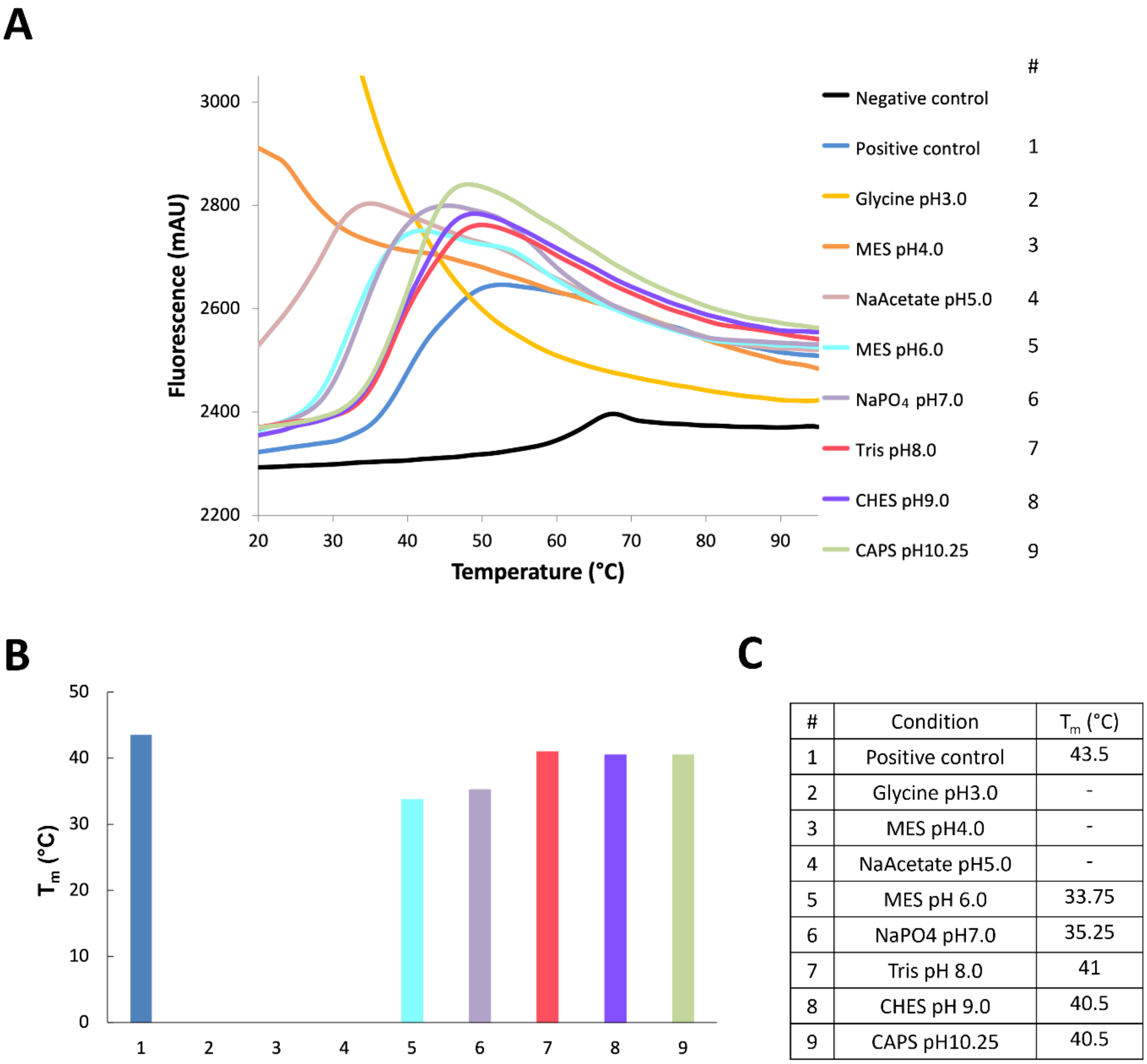
In this chapter, I describe the journey to optimize the crystallization conditions to enhance the probability of crystallization. We investigated previous crystallization conditions to have insight into RocC crystallization. TSA and DLS were used to screen optimal buffer conditions from different perspectives thermal stability and polydispersity, respectively. Many variables were tested, and techniques were performed to obtain high-quality crystals and improve the quality of the crystals. As a result of optimization, we determined apo-RocC structures and RocC in complex with the intrinsic terminator. The apo-RocC showed a highly α-helical structure similar to other previously determined FinO domains. Modified SL3 of RocR was revealed to interact through the 3' tail and base of the stem with almost all contacts occurring through the tail and 3' side of the stem. The lower stem of SL3 binds through the N-cap motif on the concave side of RocC. The 3' tail binds to the positively charged pocket. SEC-SAXS analysis of the complex revealed that crystal structure represents the conformation in solution.

4.3. Results

4.3.1. Buffer optimization for crystallization

We optimized the purification method for the designed constructs (Chapter 3.1) for structural study with little modification of the previous RocC purification method (Attaiech et al., 2016). The purified RocC was in the buffer 50 mM HEPES pH 7.3, 150 mM NaCl, 1 mM DTT (Attaiech et al., 2016). The structure of the RocC FinO domain was predicted using SWISS-MODEL (Waterhouse et al., 2018) and showed a high degree of structural similarity to the crystal structures of *E. coli* FinO (PDB ID/1DVO, RMSD = 2.09 Å) or *N. meningitidis* NMB1681 (PDB ID/3MW6, RMSD = 3.53 Å). We investigated the crystallization conditions of these two proteins to look for common themes that could provide leads for RocC crystallization. $\text{FinO}_{26\text{-}186}$ was crystallized at a final concentration at 5 mg/mL in 50 mM MES pH 6.5, 150 mM NaCl, 1 mM EDTA, 0.1% β-mercaptoethanol (v/v) buffer with a reservoir solution 12% PEG 4000, 50 mM Tris, pH 7.2 at 4 °C using hanging drop vapor diffusion method. NMB1681 was crystallized at 45 mg/ml in a solution of 20 mM Tris pH 7.5, 500 mM NaCl, 2 mM DTT, and a reservoir solution of 0.1 M Bis-tris, pH 6.5, 0.17 M $MgCl_2$, 21% (w/v) PEG3350 using sitting drop vapour diffusion method at 4°C and 18°C. A common feature of the two conditions was that both crystals were formed at lower than room temperature. Furthermore, this comparison showed that the solubility of NMB1681 was much higher than the FinO fragment that was crystallized. Our previous experience with FinO was that the full-length protein or the nearly full-length 26-186 construct was only poorly soluble and only at low temperatures. These proteins at room temperature or above rapidly precipitated when at high concentrations. In contrast, NMB1681, which is essentially just a FinO domain, was much more soluble. We had also found that more minimal FinO constructs lacking the flexible N-terminal region was, like NMB1681, much more soluble. Based on these observations, we felt that it would be critical to carefully define the minimal RocC FinO domain to obtain a protein construct that would be highly soluble and amenable to structural studies. We performed a solubility test with RocC$_{24-126}$. The test was designed to assess the solubility of RocC$_{24-126}$ in 25 mM Tris pH 8.0, 150 mM NaCl, 1 mM DTT buffer as it concentrates in a centricon until it forms white precipitate. There was no precipitation observed up to 110 mg/mL and we stopped concentrating RocC due to the shortage of protein. We failed to find the maximum solubility of RocC. However, this result was sufficient to demonstrate the highly soluble nature of the FinO domain of RocC. Literature reviews and the solubility assay led us to test various concentrations (2.5 mg/mL to 33 mg/mL) and different temperatures (4°C, 16°C, and 25°C) to crystallize RocC.

To increase the probability of crystallization, we screened for optimal buffer conditions using TSA (Ericsson et al., 2006) and DLS (Borgstahl, 2007). We used TSA to measure changes in RocC thermal stability with changes in buffer pH in a range of pH 3.0 to pH 10.25 in 50 mM buffer and 50 mM NaCl (Figure 4-2A). The results showed that RocC was denatured even at the initial temperature of 20°C at pH 3.0 and pH 4.0. MES pH 6.0 and $NaPO_4$ pH 7.0 showed similar T_m, ~33°C, which indicated that the low stability was really due to the low pH, not MES. Tris pH 8.0, CHES pH 9.0, and CAPS pH 10.25 gave a T_m ~40°C, indicating that RocC showed similar stability from pH 8.0 to pH 10.25. However, the positive control, 25 mM HEPES pH 7.3, 150 mM NaCl, displayed slightly higher T_m, 43.5 °C than second place, Tris pH 8.0. The thermal shift assay suggested that pH 7.3 HEPES and pH 8.0 Tris are an optimal buffers for RocC and these were tested in crystallization trials.



#	Condition	T_m (°C)
1	Positive control	43.5
2	Glycine pH3.0	-
3	MES pH4.0	-
4	NaAcetate pH5.0	-
5	MES pH 6.0	33.75
6	NaPO4 pH7.0	35.25
7	Tris pH 8.0	41
8	CHES pH 9.0	40.5
9	CAPS pH10.25	40.5

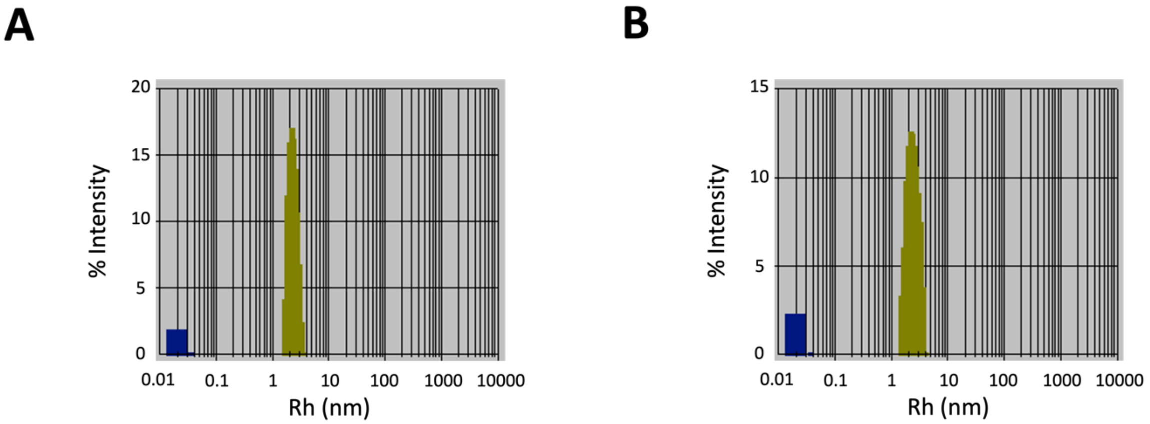
Figure 4-2. Thermal shift assay of RocC$_{24\text{-}126}$ with different pH buffers.

(A) Fluorescence signal changes of RocC$_{24\text{-}126}$ with temperature changes from 20°C to 95°C. in a range of pH 3.0 to 10.25. (B) T_m bar graph (C) T_m calculation chart and corresponding numbers.

DLS was used to screen the impact of salt concentration on protein homogeneity. DLS is a sensitive tool to estimate the size of molecules and their size distribution in the buffer. The high sensitivity of DLS toward bigger molecules allows screening diverse buffer conditions that induce aggregation or different oligomeric states. RocC$_{24\text{-}126}$ was in 25 mM Tris pH 8.0, with three different salt concentrations, 150 mM, 75 mM and 37.5 mM (Figure 4-3A-C). All results displayed two populations; a population of small hydrodynamic radius (R_H ~ 0.02 nm) representing solvent, and a second population corresponding to $RocC_{24-126}$, with over 97% of the scattering intensity. We used DLS to estimate the polydispersity of the sample. Polydispersity values of less than 30% indicates samples which are "likely to be crystallized" (Borgstahl, 2007). RocC in 150 mM, 75 mM, 37.5 mM NaCl showed 19.7, 21.3, 22.0% of polydispersity (Figure 4-3A-C), respectively, suggesting that RocC is highly monodisperse in a range of 37.5 mM to 150 mM NaCl concentration. Figure 4-3D is the repetition of Figure 4-2C the next day, indicating that the monodisperse sample is stable over time. DLS results demonstrated that RocC crystallization could be tried in lower salt conditions. In summary, buffer and salt concentration screening suggested two buffers for RocC crystallization: 30 mM NaCl, 1 mM DTT with either 25 mM HEPES pH 7.3 or 25 mM Tris pH 8.0. Solubility tests suggested RocC could be highly concentrated for crystallization under these conditions.

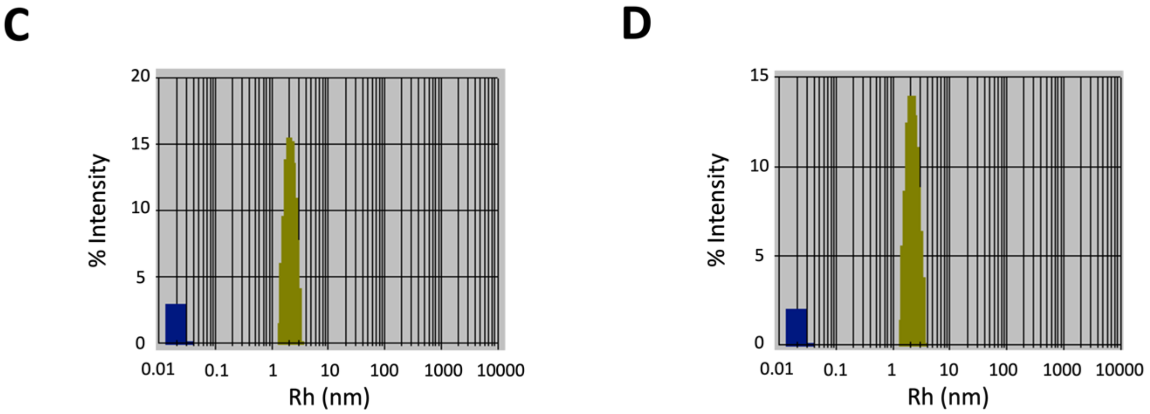
4.3.2. The strategy for apo-RocC crystallization and optimization

We attempted to crystallize 4 different apo-constructs, $RocC_{1-126,\ 14-126,\ 24-126,\ 1-137}$, each of which contain the FinO domain, and obtained crystals of truncation mutants, $RocC_{24-126}$ and $RocC_{1-126}$, suitable for structure determination. Diverse crystallization conditions were screened using commercial screening kits and the vapour diffusion method. Crystallization trials at different temperatures (4°C, 16°C, 25°C), protein concentrations (5 mg/mL to 66 mg/mL), ratio between protein and reservoir solution were all tested. The condition with 0.3 μL of 30 mg/mL of $RocC_{24-126}$ in 25 mM HEPES pH 7.3, 30 mM NaCl, 1 mM DTT mixed with 0.3 μL mother liquor gave the first diffracting crystals. The initial crystal was a thin plate, which grew radially from a single nucleation site (Figure 4-3A). The initial crystals grew up in a condition of 0.2M ammonium acetate, 0.1 M HEPES/NaOH pH 7.5, 25% PEG 3350 (TOP96 (Anatrace), F2) in sitting drops at 16°C. The optimized crystals grew in two days (Figure 4-4B) and grew bigger in three more days (Figure 4-4C). Crystals were reproducible, and bigger crystals for data collection were obtained by increasing the drop size (3-4 μL); however, they were prone to producing thin plates from a single nucleation site (Figure 4-4A-C). We tried to alter the grow pattern by lowering the temperature, setting plates with less concentrated proteins, or adding glycerol. In addition, we tried microseeding and streak seeding to obtain single crystals. Despite these efforts, crystals were prone to grow as clusters from a single nucleation site. We physically broke the crystal near the



	Radius (nm)	Polyd (nm)	% Polyd	% Int	% Mass
1st Peak	0.02	0.00	0.0	2.8	100.0
2nd Peak	2.15	0.42	19.7	97.2	0.0

	Radius (nm)	Polyd (nm)	% Polyd	% Int	% Mass
1st Peak	0.02	0.00	0.0	1.9	100.0
2nd Peak	2.23	0.48	21.3	98.1	0.0



	Radius (nm)	Polyd (nm)	% Polyd	% Int	% Mass
1st Peak	0.02	0.00	0.0	2.2	100.0
2nd Peak	2.42	0.53	22.0	97.8	0.0

	Radius (nm)	Polyd (nm)	% Polyd	% Int	% Mass
1st Peak	0.02	0.00	0.0	1.8	100.0
2nd Peak	2.32	0.40	17.2	98.2	0.0

Figure 4-3. DLS results for RocC$_{24\text{-}126}$ at different salt concentrations.

(A) DLS result of ~3.6 mg/mL RocC$_{24\text{-}126}$ in 25 mM Tris pH 8.0, 150 mM NaCl, 1 mM DTT (B) Same as condition (A), but with 75 mM NaCl. (C) Same as (A), but with 37.5 mM NaCl. (D) Repeat of (A) next day.

nucleation site while looping the crystal, yielding a single crystal that could be used in data collection. The data from this crystal was sufficient to determine the structure of the complex by molecular replacement at 2.1 Å resolution (see section 4.3.6. for details on structure determination).

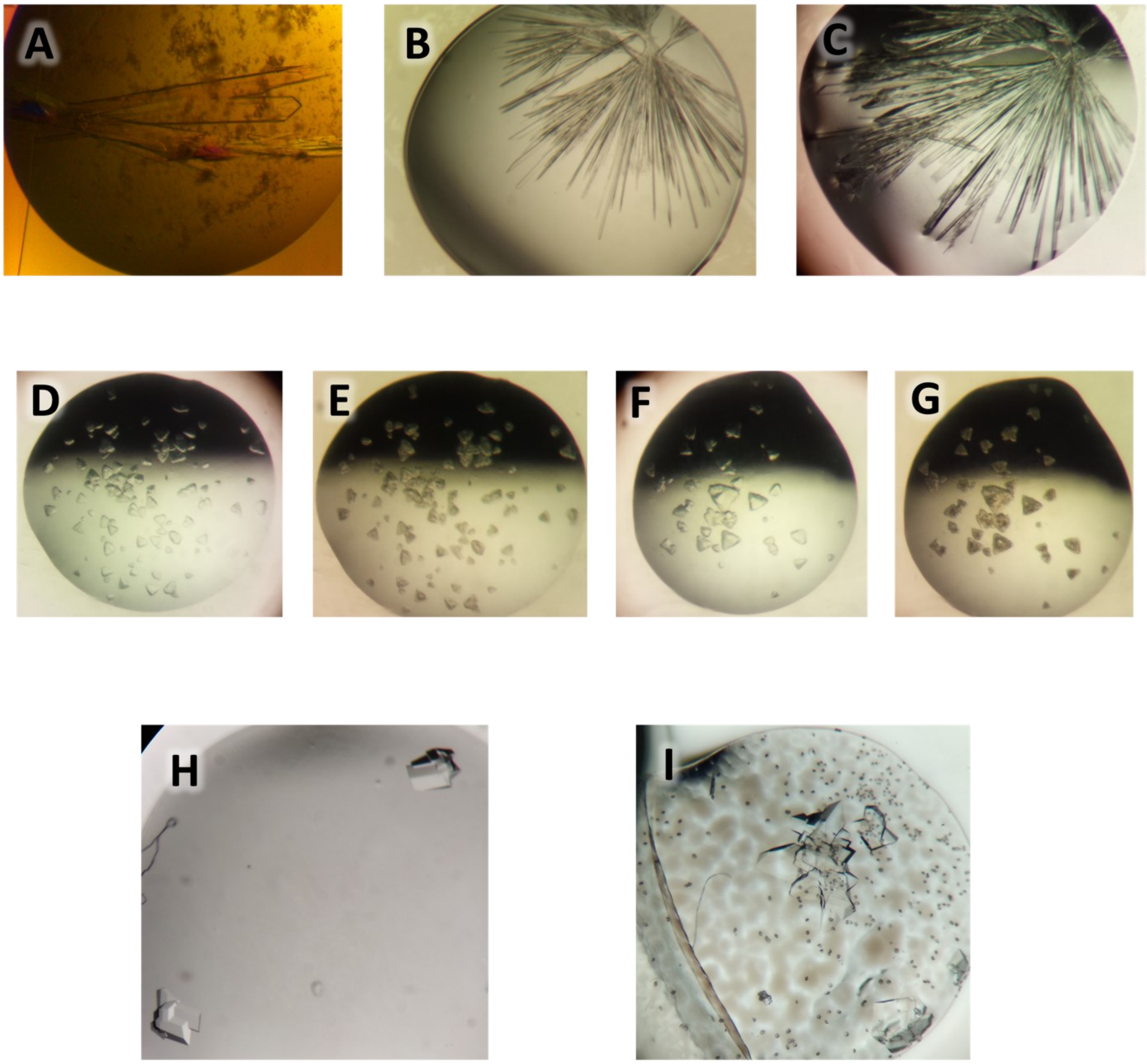
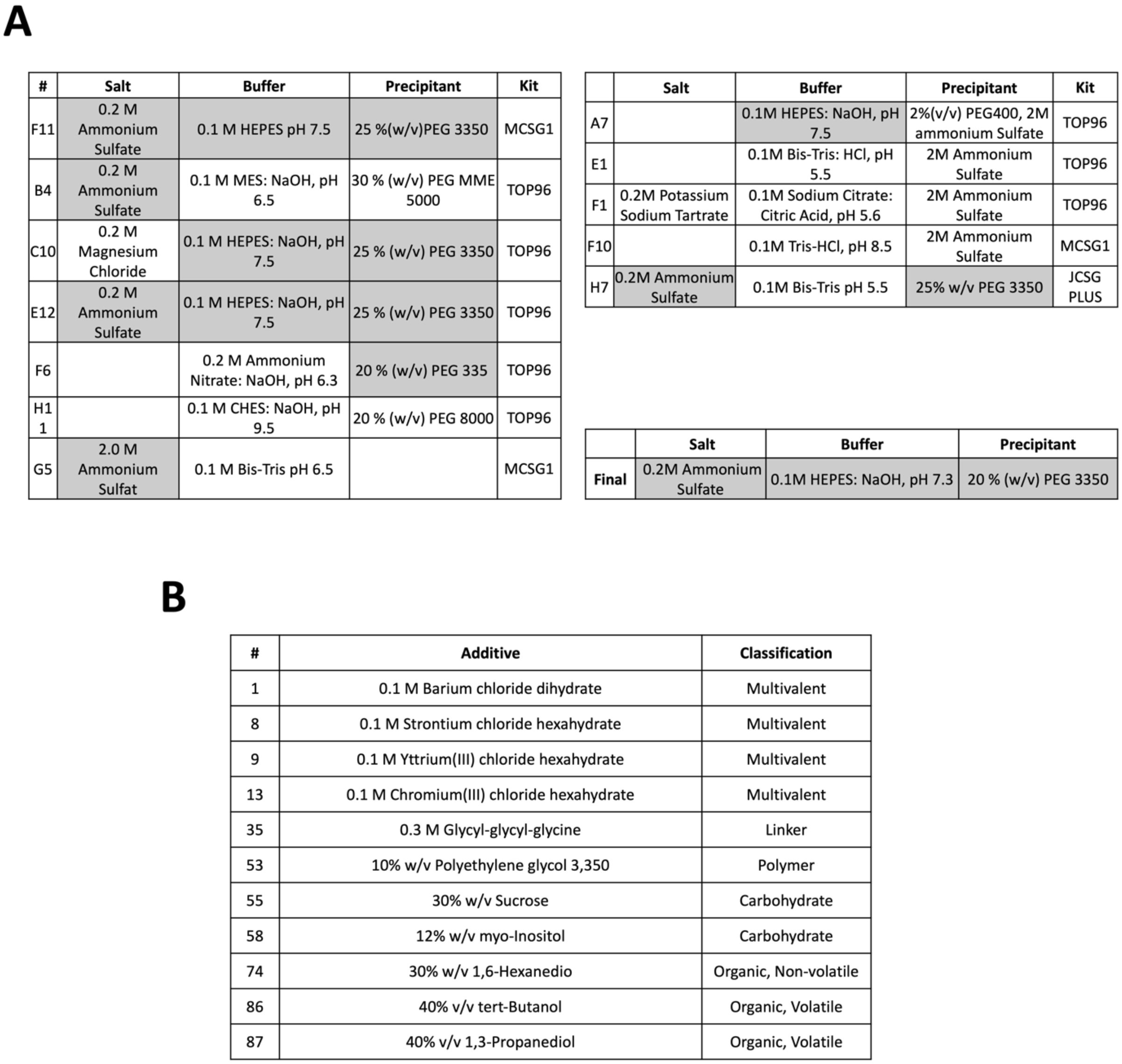


Figure 4-4. Various apo-RocC crystals.

(A) RocC$_{24\text{-}126}$ initial hits from TOP96 F2. (B) Crystals from the figure (A) were optimized in a bigger drop. Crystals grew for two days. (C) The same crystal from the figure (B) grew for five days. (D) RocC$_{1\text{-}126}$. (H) RocC$_{1\text{-}126}$ crystals with additive, 40% v/v Polypropane Glycol P400. (I) RocC$_{1\text{-}126}$ crystals with additive, 0.1 M $SrCl_2$.

We tried to obtain RocC$_{1\text{-}126}$ because the binding assay from EMSA and FP indicates defectiveness of RocC$_{24\text{-}126}$ in binding with RocR$_{SL3}$ (Chapter 3). We examined the size of the solvent channels in the RocC$_{24\text{-}126}$ structure to determine whether the RocC$_{24\text{-}126}$ lattice could accommodate the additional RocC$_{1\text{-}23}$. Unfortunately, the RocC$_{24\text{-}126}$ lattice was tightly packed, so that there was not enough space for additional 23 residues to fit in. This structural analysis suggested that the crystallization condition of RocC$_{1\text{-}126}$ should be screened from scratch. Sparse matrix screens using RocC$_{1\text{-}126}$ produced triangle-shaped crystals (Figure 4-4D, F) two days after setting the plate at 16°C in many different conditions (Table 4-1). Interestingly, one of the conditions (TOP 96 (Anatrace), E12 in Table 4-1) yielded triangular RocC$_{1\text{-}126}$ crystals in the same condition that we obtained RocC$_{24\text{-}126}$, except the salt was ammonium sulfate instead of ammonium acetate. The pictures shown in Figure 4-4 D and F were shot two days after setting the screens, and figures 4-4E and G (same drops with D and F, respectively) were shot the following day. These pictures show that the crystals quickly formed but also decayed by the third day. Despite crystal showers across the screening, crystals stopped growing in two days and quickly deteriorated. Diffraction data for these crystals were collected on our home source X-ray (Rigaku) to confirm whether the crystals were from protein or salt. Long exposure images indicated weak diffraction from the triangular crystals (Data not shown), suggesting the triangular crystals are protein. We tried microseeding and streak seeding, as well as growth temperature and protein concentration optimization, to improve these crystals. A comparison of the crystallization conditions of the initial hits suggested that they share similar conditions such as 0.2 M $(NH_4)_2SO_4$ and 0.1 M HEPES pH 7.5. Around 20-25% PEG 3350 was a common precipitant (grey coloured in Table 4-1A). The crystallization condition analysis helped determine the reservoir condition, and it turned out to be the same as one of the screening conditions, MCSG1 (Anatrace, F11), which consists of 0.2 M Ammonium sulfate, 0.1 M HEPES pH 7.5, 25% PEG 3350. We tried additive screens (Hampton HR2-428) using 0.2M Ammonium sulfate, 0.1M HEPES pH 7.5, 25% PEG 3350 as a reservoir solution to find additives to give the crystals stability. The crystal plates with additive screens were set in low temperatures (16°C to 4°C) to increase the protein stability in the course of being crystallized. This optimization strategy gave us sharp-edged crystals with multiple additives in the crystal plates sitting at 4°C. Even though RocC$_{1\text{-}126}$ crystals grew up in various additives, they had the same shape, indicating they were in the same space group. In conclusion, we obtained high-quality RocC$_{1\text{-}126}$ crystals for data collection from the various additives, from multivalent ions to organic/volatile solvent (Figure 4-4H, I and table 4-1B).



#	Salt	Buffer	Precipitant	Kit
F11	0.2 M Ammonium Sulfate	0.1 M HEPES pH 7.5	25 %(w/v)PEG 3350	MCSG1
B4	0.2 M Ammonium Sulfate	0.1 M MES: NaOH, pH 6.5	30 % (w/v) PEG MME 5000	TOP96
C10	0.2 M Magnesium Chloride	0.1 M HEPES: NaOH, pH 7.5	25 % (w/v) PEG 3350	TOP96
E12	0.2 M Ammonium Sulfate	0.1 M HEPES: NaOH, pH 7.5	25 % (w/v) PEG 3350	TOP96
F6		0.2 M Ammonium Nitrate: NaOH, pH 6.3	20 % (w/v) PEG 335	TOP96
H11		0.1 M CHES: NaOH, pH 9.5	20 % (w/v) PEG 8000	TOP96
G5	2.0 M Ammonium Sulfat	0.1 M Bis-Tris pH 6.5		MCSG1

	Salt	Buffer	Precipitant	Kit
A7		0.1M HEPES: NaOH, pH 7.5	2%(v/v) PEG400, 2M ammonium Sulfate	TOP96
E1		0.1M Bis-Tris: HCl, pH 5.5	2M Ammonium Sulfate	TOP96
F1	0.2M Potassium Sodium Tartrate	0.1M Sodium Citrate: Citric Acid, pH 5.6	2M Ammonium Sulfate	TOP96
F10		0.1M Tris-HCl, pH 8.5	2M Ammonium Sulfate	MCSG1
H7	0.2M Ammonium Sulfate	0.1M Bis-Tris pH 5.5	25% w/v PEG 3350	JCSG PLUS

	Salt	Buffer	Precipitant
Final	0.2M Ammonium Sulfate	0.1M HEPES: NaOH, pH 7.3	20 % (w/v) PEG 3350

#	Additive	Classification
1	0.1 M Barium chloride dihydrate	Multivalent
8	0.1 M Strontium chloride hexahydrate	Multivalent
9	0.1 M Yttrium(III) chloride hexahydrate	Multivalent
13	0.1 M Chromium(III) chloride hexahydrate	Multivalent
35	0.3 M Glycyl-glycyl-glycine	Linker
53	10% w/v Polyethylene glycol 3,350	Polymer
55	30% w/v Sucrose	Carbohydrate
58	12% w/v myo-Inositol	Carbohydrate
74	30% w/v 1,6-Hexanedio	Organic, Non-volatile
86	40% v/v tert-Butanol	Organic, Volatile
87	40% v/v 1,3-Propanediol	Organic, Volatile

Table 4-1. Crystallization conditions of RocC$_{1\text{-}126}$. (A) Initial conditions where triangular-shape crystals grew (B) Additive screens where the improved quality of RocC$_{1\text{-}126}$ grew.

4.3.3. The development of the RNA purification protocol and optimization

In vitro transcription is a well-established method to produce RNA using T7 RNA polymerase and DNA template. Denaturing gel electrophoresis followed by phenol/chloroform extraction and ethanol precipitation is a conventional way to purify a large amount of target RNA from *in vitro* transcription reactions. However, this method is time-consuming and purity can be an issue for certain RNAs. We overcame these issues by optimizing the RNA purification method using anion exchange and gel filtration chromatography (Easton et al., 2010).

The first goal was to test the separation capacity of anion exchange chromatography (Resource Q, GE healthcare) and obtain a homogenous population of target RNA. The mixture of 3 mL *in vitro* transcription was loaded into Resource Q (Figure 4-5A). A large peak from the RNA was eluted beginning at about 45% buffer B (i.e. ~0.45 M NaCl). We ran samples from the eluted peak on native gels and stained with an intercalating agent, SYBR safe (ThermoFisher, Figure 4-5B) to characterize recovery and purity. Despite the fractions all coming from one peak, lanes 3 to 6 did not show any bands, but lanes 7 to 10 showed stained RNA bands. The black arrow possibly corresponds to a monomer of the intact RNA and the hollow arrow could be a dimer of the same RNA (Figure 4-5A). To test this idea, samples in lanes 7-9 were mixed and annealed at 95°C for 2 minutes, and the annealed mixture was snapped cooled in the ice. Annealed RNA was run in lane 1, showing that most of the bands from the hollow arrow converged into the black arrow, consistent with the idea that the species indicated by the hollow arrow are actually a dimer species (lane 1 in Figure 4-5B). To increase the separation capacity of the resource Q and spread this heterogenous peak, we adjusted to a shallower salt gradient in the range between 35%-60% where the RNA elutes (Figure 4-5C). This adjustment increased the separation capacity of the resource Q so that the monomer RNA could be separated from both aborted or self-dimerization species (Figure 4-5C, D). Purified RNAs from anion exchange chromatography were loaded onto gel filtration to exchange high salt buffer for the desired buffer for crystallization and confirm the monomeric state (Figure 4-6C, D).

We focused on increasing the yield for crystallization after establishing the protocol to purify high-quality RNAs. Highly soluble biomacromolecules often require higher concentration than molecules with lower solubility to grow crystals and this can be especially true for nucleic acids. We tried to obtain a high yield of the RNA from *in vitro* transcription by optimizing the incubation time and component concentration. We found that a 3 hours transcription was sufficient to maximize yield (Table 4-2). Different concentrations of DNA template, $MgCl_2$, and T7 RNA polymerase, were tested with 3 hours incubations to find conditions with increased yield. These studies helped us uncover conditions that significantly improved yield however, we also found that

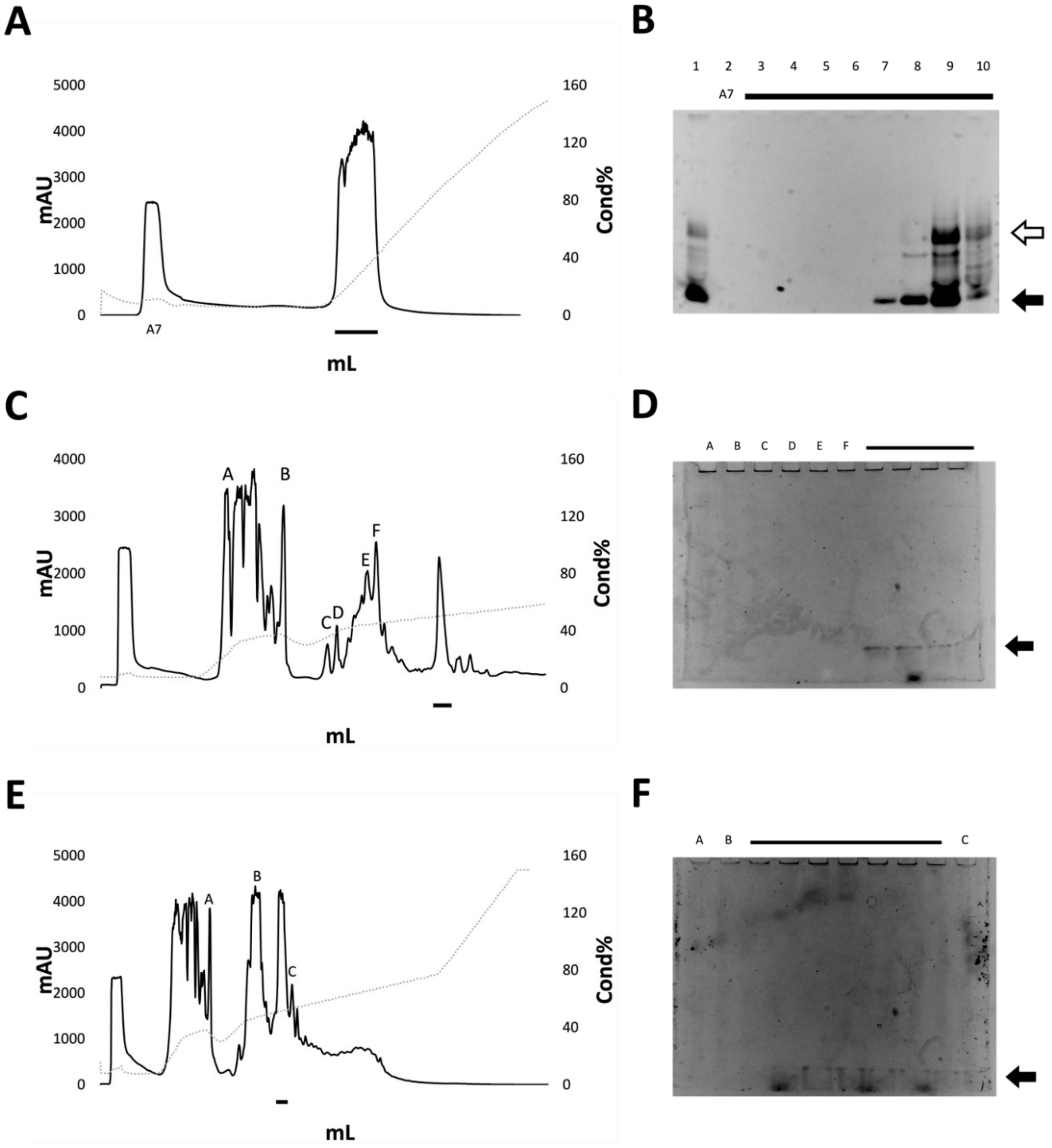


Figure 4-5. RNA purification using anion exchange chromatography from *in vitro* transcription.

(A) Chromatogram of anion exchange chromatography of RocR$_{\text{9bp-tet}}$. The solid line indicates the absorbance at 280 nm, and dotted line indicates the conductivity. (B) 10% Tris-glycine native gel of (A). The black bar in (A) corresponding to the grey bar on (B). The black arrow indicates monomer. The hollow arrow indicates oligomeric RNA. (C) Chromatogram of anion exchange chromatography of RocR$_{\text{9bp-tet}}$ with a gradual salt gradient. (D) 10% Tris-glycine native gel of (C). (E) Chromatogram of anion exchange chromatography of RocR$_{\text{SL3}}$ with a gradual salt gradient. (F) 10% Tris-glycine native gel of (E).

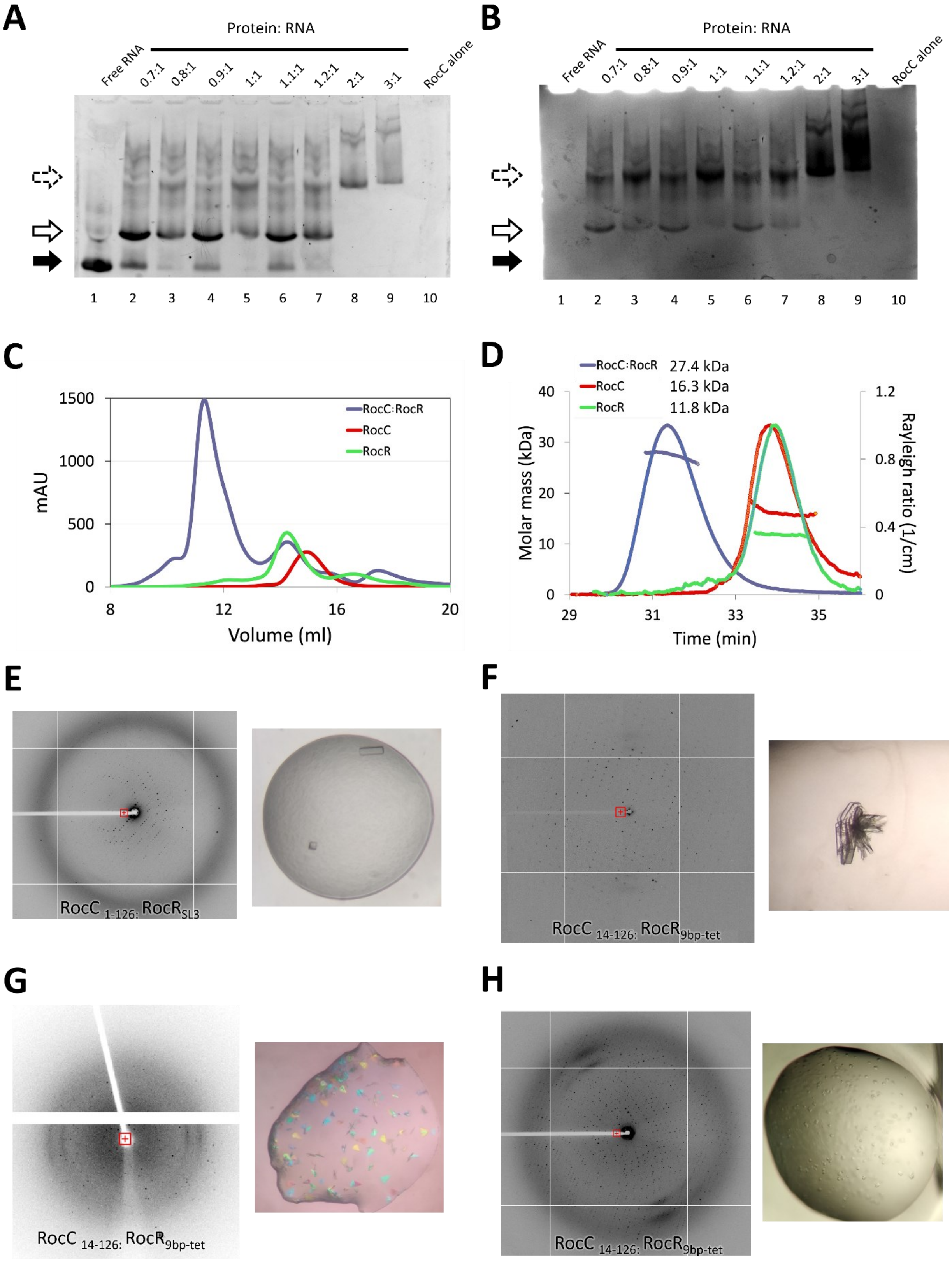

A
Protein: RNA
Free RNA
0.7:1
0.8:1
0.9:1
1:1
1.1:1
1.2:1
2:1
3:1
RocC alone
1
2
3
4
5
6
7
8
9
10
B
Protein: RNA
Free RNA
0.7:1
0.8:1
0.9:1
1:1
1.1:1
1.2:1
2:1
3:1
RocC alone
1
2
3
4
5
6
7
8
9
10
C
mAU
1500
1000
500
0
8
12
16
20
Volume (ml)
RocC:RocR
RocC
RocR
D
RocC:RocR 27.4 kDa
RocC 16.3 kDa
RocR 11.8 kDa
Molar mass (kDa)
40
30
20
10
0
29
31
33
35
Time (min)
Rayleigh ratio (1/cm)
1.2
0.8
0.4
0
E
RocC 1-126: RocR SL3
F
RocC 14-126: RocR 9bp-tet
G
RocC 14-126: RocR 9bp-tet
H
RocC 14-126: RocR 9bp-tet

Figure 4-6. The optimization and crystallization of RocC:RocR complex.

Different ratios of protein:RNA were incubated and run on a 10% Tris-glycine gel. (A) the result of staining with SYBR safe (B) the result of staining with Coomassie blue. The black arrow indicates free RNA. A solid, hollow arrow indicates a one-to-one protein:RNA complex. Dotted, hollow arrow indicates oligomeric protein/RNA complex. (C) RocC:RocR complex separation from excess RNA. Blue is the chromatogram of RocC incubated with an excess of RNA. Red is RocC only. Green is RocR only. (D) Result of SEC-MALS. The complex formed a stable 1:1 complex. (E-H) Crystals from protein:RNA complex and its diffraction.

the optimal transcription conditions were highly dependent of the particular template being transcribed.

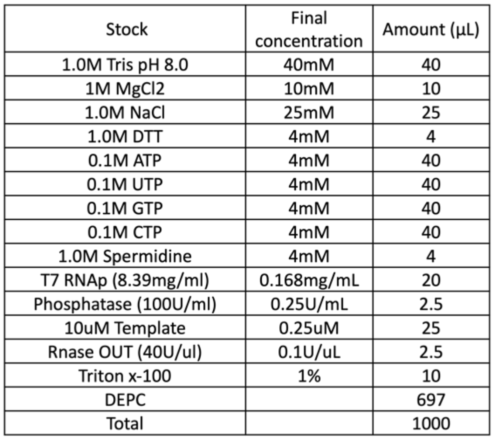
Together, the ion exchange chromatography combined with gel filtration provides high purity of RNAs. Testing the effect of different components in the *in vitro* transcription with the established purification method allowed us to obtain increased amounts of homogenous RNAs for crystallization.

4.3.4. The strategies to obtain the crystal structure of the protein:RNA complex

We attempted to crystallize and determine the crystal structure of the RocC:RocR complex to understand the molecular mechanism of RocR recognition by RocC. Here are the strategies we used to obtain the crystal structure of RocC in complex with RNA.

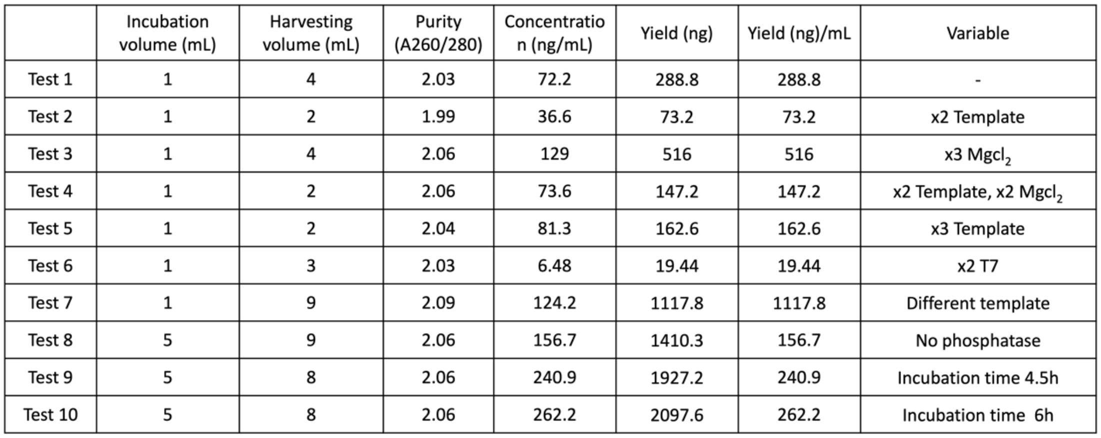
In chapter 3, we determined the minimal RocR RNA necessary to bind RocC with high affinty. These studies showed a minimum of a 5 base-pair stem and a 5 nucleotide 3’ tail is essential for RocC recognition. We did try to soak in small RNAs into our apo-RocC crystals. We attempted to soak in a single UMP with the thought that this might reveal binding of the 3’ terminal nucleotide. Analysis of the crystal packing environment suggested this small molecule might potentially be able to diffuse into the crystals. Pre-grown RocC$_{24\text{-}126}$ crystals were soaked in up to 500 mM UMP for a maximum of 24 hours at 16°C. We also tried co-crystallization of RocC$_{24\text{-}126}$ with UMP. 66 mg/mL of RocC$_{24\text{-}126}$ was incubated with 50-400 mM of UMP prior to setting crystal plates (final UMP concentration in the drop was 25-200 mM). Thin plates were grown in the co-

A



Stock	Final concentration	Amount (µL)
1.0M Tris pH 8.0	40mM	40
1M MgCl2	10mM	10
1.0M NaCl	25mM	25
1.0M DTT	4mM	4
0.1M ATP	4mM	40
0.1M UTP	4mM	40
0.1M GTP	4mM	40
0.1M CTP	4mM	40
1.0M Spermidine	4mM	4
T7 RNAp (8.39mg/ml)	0.168mg/mL	20
Phosphatase (100U/ml)	0.25U/mL	2.5
10uM Template	0.25uM	25
Rnase OUT (40U/ul)	0.1U/uL	2.5
Triton x-100	1%	10
DEPC		697
Total		1000

B



	Incubation volume (mL)	Harvesting volume (mL)	Purity (A260/280)	Concentration (ng/mL)	Yield (ng)	Yield (ng)/mL	Variable
Test 1	1	4	2.03	72.2	288.8	288.8	-
Test 2	1	2	1.99	36.6	73.2	73.2	x2 Template
Test 3	1	4	2.06	129	516	516	x3 Mgcl$_2$
Test 4	1	2	2.06	73.6	147.2	147.2	x2 Template, x2 Mgcl$_2$
Test 5	1	2	2.04	81.3	162.6	162.6	x3 Template
Test 6	1	3	2.03	6.48	19.44	19.44	x2 T7
Test 7	1	9	2.09	124.2	1117.8	1117.8	Different template
Test 8	5	9	2.06	156.7	1410.3	156.7	No phosphatase
Test 9	5	8	2.06	240.9	1927.2	240.9	Incubation time 4.5h
Test 10	5	8	2.06	262.2	2097.6	262.2	Incubation time 6h

Table 4-2. *In vitro* transcription conditions and yield test. (A) The initial *in vitro* transcription. (B) The results of *in vitro* transcription yield test.

crystallization trials at 25-100 mM UMP concentration. However, over 100 mM UMP crystals did not form. Unfortunately, we could not observe positive density for UMP in either the soaking or co-crystallization trials.

The conventional method to obtain crystals of protein:RNA complexes is co-crystallization. Excess RNA is incubated with protein prior to setting crystal plates. We investigated the ratio to form the most homogenous protein:RNA complex by native gel electrophoresis. The different protein:RNA complex ratios were incubated for 30 minutes on ice, and samples were loaded onto the nondenaturing gel. The native gel was stained with SYBR safe to visualize RNA, followed by Coomassie blue staining to visualize protein. SYBR safe stained gel showed a potential one-to-one complex in lanes 2 to 7 (Figure 4-6A). Free RNAs were observed from lanes 2 to 6, indicating that a non-bound RNA population exists at a 1.1:1 ratio. Free RNAs started to disappear in lanes 7, 8, and 9, suggesting all RNAs form protein:RNA complex from a 1.2:1 protein:RNA ratio. Hollow dotted arrows on SYBR safe stained gel indicate potential oligomeric states of the protein:RNA complex. The oligomeric complex appeared more evident after the native gel was stained with Coomassie blue (Figure 4-6B). Lane 2-7 showed populations of at least two complexes. Lane 8 and 9 showed only the band for the oligomeric states. Thus the test of different incubation ratios showed the formation of heterogeneous complexes. Observing the heterogeneity in the incubated mixture suggested the necessity for separating the one-to-one complex from others.

We tested whether the 1:1 complex in the incubated mixture can be purified through size exclusion chromatography to obtain a homogenous population (Figure 4-6C). The RocC:RocR mixture chromatogram showed multiple peaks, suggesting multiple species that could be separated based on size. Comparison to free RocC and RocR controls suggested that the major peak obtained in the complex sample indeed corresponded to a RocC:RocR complex. SEC-MALS of this purified peak revealed that this species indeed corresponds to the 1:1 RocC:RocR complex (Figure 4-6D). Furthermore, the different sets of complexes involving either $\mathrm{RocC}_{1\text{-}126}$ and $\mathrm{RocC}_{14\text{-}126}$, and either $\mathrm{RocR}_{\mathrm{SL3}}$ or $\mathrm{RocR}_{9\mathrm{bp\text{-}tet}}$, were created and also shown to form 1:1 complexes by SEC-MALS (Figure 3-7C, D, E).

4.3.5. Crystallization of RocC in complex with RNA

Extensive sparse matrix screens were conducted with protein/RNA complexes, purified with the established purification protocol. The screens were carried out at different complex concentrations and crystallization temperatures, and we eventually succeeded in growing diffraction quality crystals (Figure 4-5E, F, G, H).

$RocC_{1\text{-}126}$:$RocR_{SL3}$ crystals were grown in 0.09 M NPS, 0.1 M buffer System 2 7.5 30 % v/v precipitant Mix 2 (Morpheus, Molecular dimension, C6) at 16°C (Figure 4-5E). Buffer system 2 is a mixture of 1.0 M sodium HEPES and 1.0 M MOPS adjusted to pH 7.5 and precipitant mix is a mixture of 40% v/v ethylene glycol and 20% w/v PEG8000. Detailed calculation of buffer system 2 and precipitant mix 2 can be found in this reference (Gorrec, 2009). This condition produced a rectangular parallelepiped shape crystal. It was a single crystal that diffracted to 4.5 Å. Unfortunately, ice formation during crystal screening prevented further data collection, and we could not reproduce these crystals for further study.

$RocC_{14\text{-}126}$:$RocR_{9bp\text{-}tet}$ complex was crystallized in three different conditions at 25°C. Triangular thin crystals (Figure 4-5G) grew in 0.2 M $CaCl_2$ 0.1 M HEPES/NaOH, pH 7.5 28 % (v/v) PPG P400 (MCSG1, Anatrace, A10). The clustered microcrystals were optimized, to grow single and bigger crystals. The optimized, single crystals were very fragile and were difficult to loop for freezing. The irregular diffraction of the crystal implied that it is loosely packed and required further optimization. We optimized the crystal by changing crystallization temperature and complex concentration to obtain a single, high-quality crystals. In addition, additive screens were performed from the experience that produced more stable $RocC_{1\text{-}126}$ crystals (Figure 4-4D-I). In the end, we succeeded to obtain single crystals, however, diffraction quality was not visibly improved (Figure 4-5G).

Another crystal form was obtained from $RocC_{14\text{-}126}$:$RocR_{9bp\text{-}tet}$, which grew as stacked thin plates from solutions of 0.2 M $CaCl_2$ 0.1 M Bis-Tris/HCl, pH 6.5 45 % (v/v) MPD (MCSG2, Anatrace, H1). Grid screening improved the size of the crystals, but they still grew as layered thin plates. The efforts to make separated and thicker crystals by controlling temperature, protein concentration and additive screening were not successful and instead we decided to work with the clustered thin plates. A single crystal was separated physically, and initial images with the incident X-ray beam orthogonal to the plate revealed diffraction to 2.5 Å resolution (Figure 4-6F). However, rotation of the crystal in the beam revealed disorder and only weak diffraction in other orientations which was not suitable for data collection.

We also obtained diffracting crystals from $RocC_{14\text{-}126}$:$RocR_{9bp\text{-}tet}$ in 0.2 M Li_2SO_4 0.1 M Tris/HCl, pH 8.5 30 % (w/v) PEG 4000 (Top96, Anatrace, A10). We obtained these crystals from a sparse matrix screen however we failed to reproduce them. Fortunately, we could collect a data set using a microbeam (20 μm x 20 μm) at ALS beamline (BEAMLINE 8.2.2 with ADSC QUANTUM 315r detector). Data collection revealed the complex crystallizes in the P2 space group and diffracts to 3.2 Å resolution. The crystal contained weak diffraction from a secondary lattice. We could separate the primary crystal diffraction from second crystal by increasing the intensity threshold for reflections used during the integration process in Denzo (Chapter 2). This data set was used to eventually determine the structure of the RocC:RocR complex.

4.3.6. Apo-crystal structure determination and structural analysis

We determined two apo-protein crystal structures, $\text{RocC}_{24\text{-}126}$ and $\text{RocC}_{1\text{-}126}$, using MR at 2.1 Å and 2.0 Å, respectively (Chapter 2). $\text{RocC}_{24\text{-}126}$ constitutes a proteolytically stable core, and the crystal structure reveals a compact and largely α-helical structure (Figure 4-7B). $\text{RocC}_{1\text{-}126}$ has the same core structure however the 1-23 segment adopts a flexible extended structure which is partially visualized in 2 of the 9 protomers in the asymmetric unit, suggesting that residues $\text{RocC}_{1\text{-}7}$ adopts α-helical conformation (Figure 4-7A). The core $\text{region}_{24\text{-}126}$ is relatively rigid, with an RMSD of 0.29 Å between $\text{RocC}_{1\text{-}126}$ and $\text{RocC}_{1\text{-}126}$. The only region of significant flexibility is in a single loop linking α2 and α4 (Figure 4-7B). The overall structure is very similar to FinO, with the shape of a right-handed fist with an extended index finger (Figure 4-7B). Superimposing the crystal structures of RocC and FinO results in a Cα-RMSD of 0.78 Å, indicating high similarity in structure (Figure 4-7C).

Unlike the well-conserved core FinO domain, density for the region N-terminal to residue 14 was weak, and this region could not be modelled in all the protomers. Density for α1 ($\text{RocC}_{1\text{-}7}$) was only observed in 2 protomers of the 9 protomers in $\text{RocC}_{1\text{-}126}$ asymmetric unit and density for the linker between α1 and α2 was not observed in any of the protomers. Overlay of FinO, RocC and 6 copies of NMB1681 crystal structures reveals that while the core FinO domains are structurally well conserved, regions N- and C-terminal to this domain are generally not well packed against the rest of the structure and are likely flexible (Figure 3-1B). This was also observed in the comparison of the 10 final NMR structures of Lpp1663 in *L. pneumophila,* which aligned well within the FinO domain (19-121, RMSD = 1.77 Å). However, the N- and C-terminal regions could not be accurately modeled due to a lack of NMR restraints.

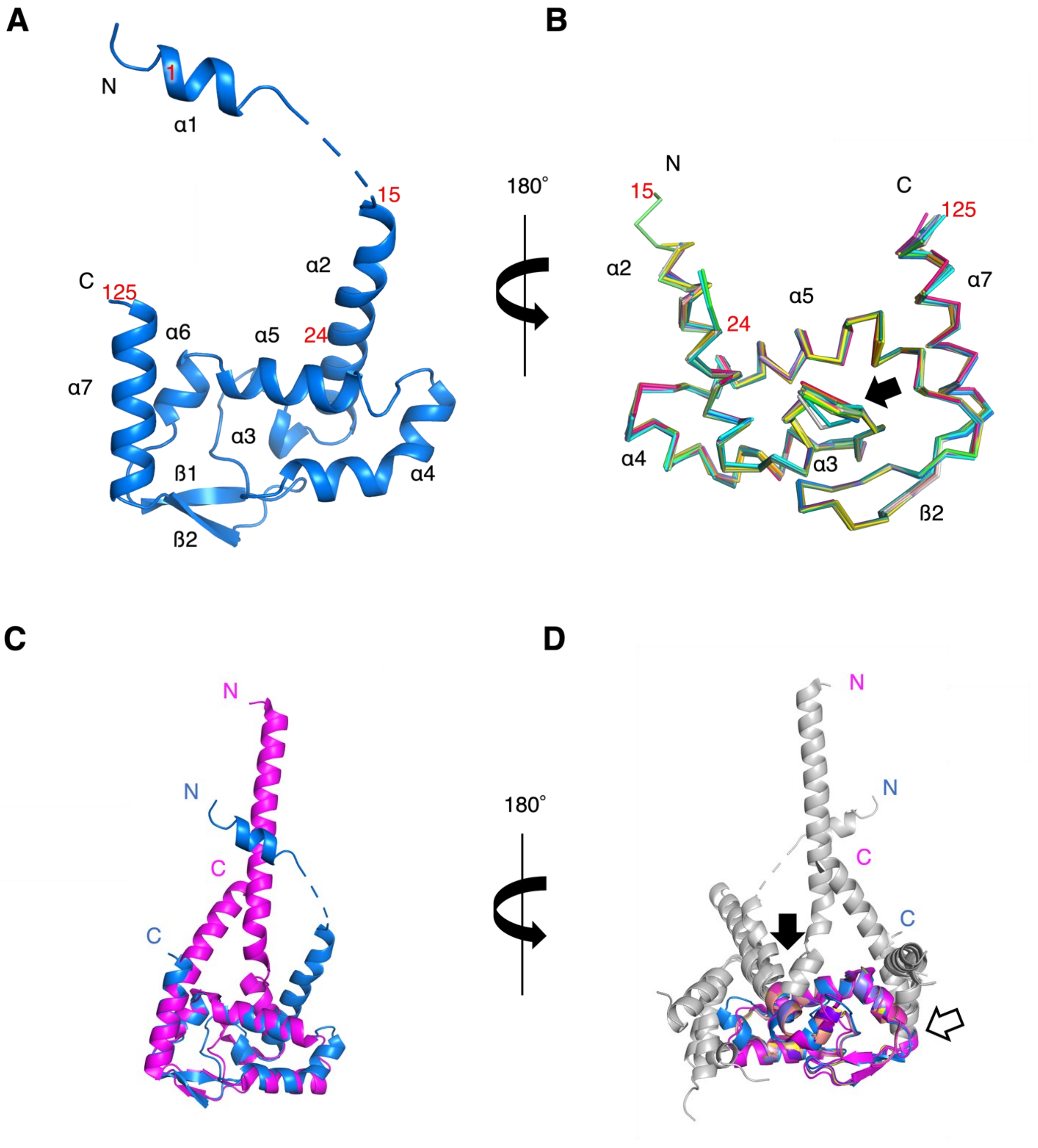


Figure 4-7. Apo-RocC structure.

Two apo-protein crystal structures, RocC$_{24\text{-}126}$ and RocC$_{1\text{-}126,}$ were determined at 2.1 Å and 2.0 Å, respectively (see Chapter 2, Table 4-3). The RocC$_{1\text{-}126}$ crystal contains 9 protomers in the asymmetric unit, while the RocC$_{24\text{-}126}$ crystal contains a single protomer in the asymmetric unit. RocC$_{24\text{-}126}$ constitutes a proteolytically stable core, and the crystal structure reveals a compact and well-folded structure. RocC$_{1\text{-}126}$ (A) has the same core structure, however, the 1-23 segment adopts a flexible extended structure which is partially visualized in 2 of the 9 protomers in the asymmetric unit, suggesting that residues 1-7 have the potential to adopt a helical conformation. The core region $\text{RocC}_{24\text{-}126}$ is quite rigid, with an rmsd of 0.29 Å between $\text{RocC}_{1\text{-}126}$ and $\text{RocC}_{24\text{-}126}$ structures. The only region of significant flexibility is in a single loop linking α2 and α4. (B) Cartoon view of apo-RocC 1-126 with secondary structure elements and key residue positions labeled. (B) Cα traces of the aligned 9 $\text{RocC}_{1\text{-}126}$ protomers and single $\text{RocC}_{24\text{-}126}$ protomers, as well as the 10 copies of $\text{RocC}_{24\text{-}126}$ in the RocC:RocR crystallographic asymmetric unit. The view is rotated 180° relative to (a), and the black arrow indicates the conformationally flexible loop. (C) Superimposition of $\text{RocC}_{1\text{-}126}$ and FinO structure (PDB ID: 1DVO) with an rmsd of 0.78 Å.

4.3.7. Structural analysis of the RocC:RocR complex

We were able to determine the structure of $\text{RocC}_{14\text{-}126}$ bound to $\text{RocR}_{9\text{bp-tet}}$ using the $P2_12_12$ crystals described above and the final structure was refined to 3.2 Å (Figure 4-8A). The asymmetric unit contains six copies of apo-$\text{RocC}_{14\text{-}126}$ and four copies of the protein-RNA complex. Overall, the hairpin-tail RNA binds to one side of the ProQ/FinO domain, mainly through interactions between the protein and the RNA backbone.

NCS averaged maps revealed electron density of sufficient quality to model the entire hairpin-tail RNA (Figure 4-9). As predicted (Figure 4-8D), the 9 base-pairs stem is fully base-paired, however, the 5' single nucleotide, U_1, forms an additional mismatch base-pair with 3' U_{24} and this base pairing is confirmed by analysis of ^{1}H NMR spectra of the free and RocC-bound $\text{RocR}_{9\text{bp-tet}}$ (Figure 4-10). The hairpin adopts the expected A-form helical geometry and the 5'-$U_{11}U_{12}C_{13}G_{14}$-3' tetraloop (Bottaro & Lindorff-Larsen, 2017), adopts the expected structure. The 4-nucleotide single-stranded 3' tail (5'-$U_{25}U_{26}C_{27}U_{28}$-3') adopts a hook-like structure. U_{25} maintains an A-form geometry, however U_{26} and C_{27} bend away, unstacking from U_{25} and instead stacking upon each other.

The 3' terminal nucleotide, U_{28} bends back such that its 3'-OH hydrogen bonds with the penultimate phosphate linking U_{26} and C_{27} (Figure 4-8A, B, Figure 4-11A). The terminal U_{28} is bound within a well-conserved and structurally rigid pocket that is formed between α5 and a β-turn at the N-terminus of α4. The pocket contains Tyr87 and Arg97, which are both among the most highly conserved residues in the ProQ/FinO domain family and together recognize the phosphate of U_{28} (Figure 4-8B, C). The β-turn-α-helix motif packs against the minor groove face of U_{28}, hydrogen bonding with the U_{28} 2' and 3' hydroxyl groups. Gly52, which stabilizes the turn,

	RocC$_{24\text{-}126}$ (PDB ID: 7RGS)	**RocC$_{1\text{-}126}$ (PDB ID: 7RGT)**	**RocC$_{14\text{-}126:}$ RocR$_{9bp\text{-}tet}$ (PDB ID: 7RGU)**
Data collection			
Space group	$P2_12_12$	C2	$P2_12_12_1$
Cell dimensions			
a, b, c (Å)	55.56, 81.77, 49.42	139.01, 100.30, 100.72	87.81, 135.94, 156.76
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 107.57, 90.00	90.00, 90.00, 90.00
Resolution	50.00-2.10 (2.14-2.10)	50.00-2.02 (2.05-2.02)	50.00-3.20 (3.31-3.20)
R_{meas}	0.134 (0.565)	0.123 (1.186)	0.334 (0.883)
I/σ (I)	35.70 (7.33)	17.43 (1.50)	4.98 (1.67)
Completeness (%)	97.9 (91.8)	99.7 (95.6)	91.4 (92.0)
Redundancy	36.5 (28.7)	6.2 (4.7)	4.4 (4.5)
Refinement			
Resolution range (Å)	49.42-2.10 (2.31-2.10)	45.30-2.02 (2.10-2.02)	48.77-3.20 (3.31-3.20)
No. reflection	13,164 (1,257)	94,408 (9,025)	28,975 (2,864)
R_{work}/R_{free}	0.161 (0.170)/0.238 (0.223)	0.193 (0.281)/0.213 (0.308)	0.219 (0.243)/0.271 (0.333)
No. atoms	2,087	8,197	20,397
Protein	1,724	7,669	17,113
RNA/ ligand	-	45	3,284
Water	363	483	-
B-factors			
Protein	25.1	67.7	44.8
RNA/ ligand	-	65.2	71.0
Water	33.7	52.2	-
r.m.s deviations			
Bond lengths (Å)	0.006	0.005	0.002
Bond angles (°)	0.74	0.71	0.59

Table 4-3. Data collection and statistics of RocC structure determination.

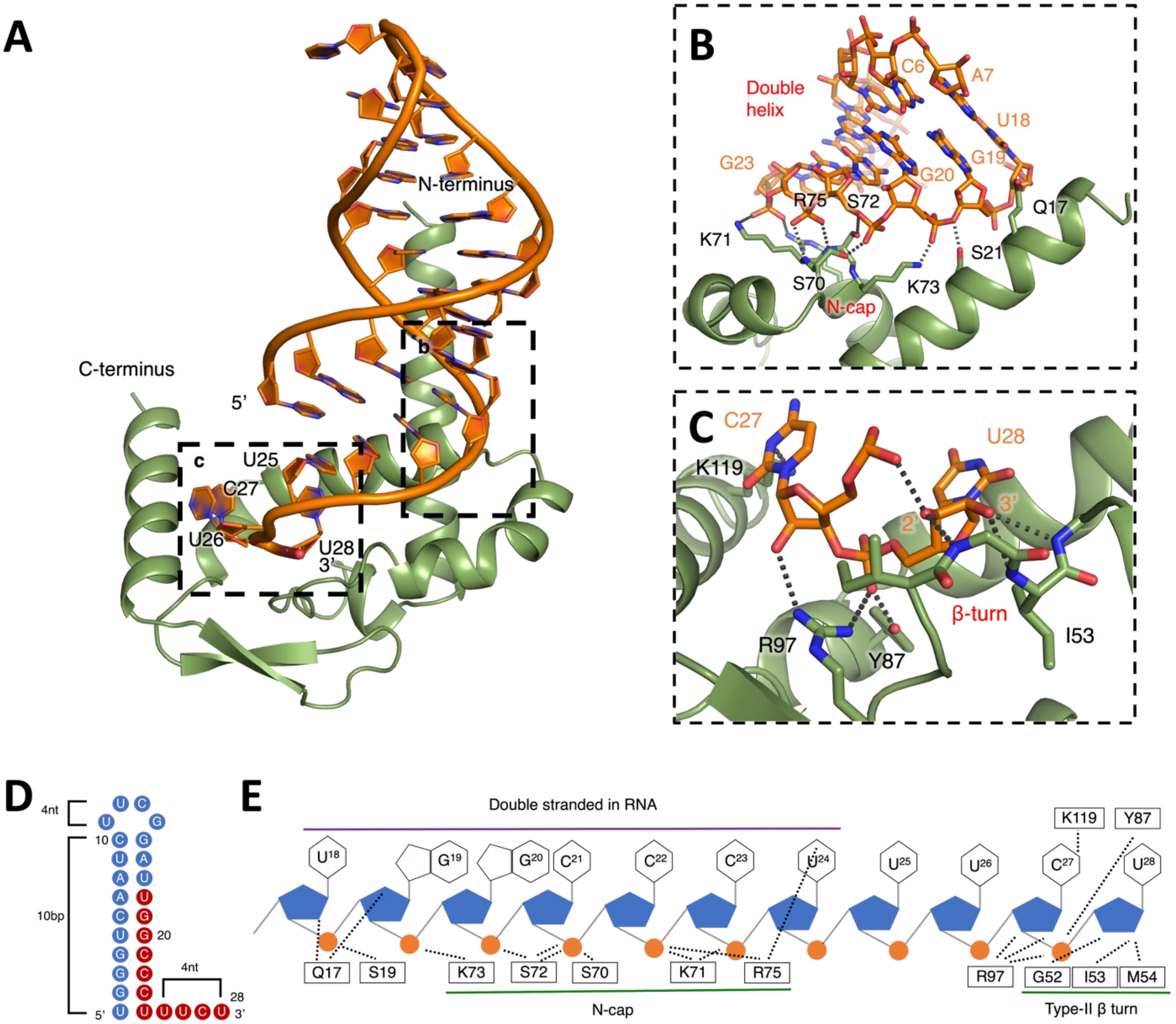


Figure 4-8. Structural analysis of the FinO domain of RocC bound to an RocR$_{SL3}$variant.

(A) Crystal structure of RocC$_{14\text{-}126}$:RocR$_{9bp\text{-}tet}$ complex. Dotted boxes indicate the main interactions between protein-RNA. (B) and (C) are zoomed-in views from figure 4-8A. Dotted lines indicate hydrogen bonding between protein and RNA. (D) Schematic diagram of the RocR$_{9bp\text{-}tet}$ variant, which is crystallized with RocC. The red circles indicate the nucleotides in direct contact with the protein. (E) Schematic diagram of RocC:RocR interactions. The purple line indicates the region of double stranded RNA structure; green lines indicate protein motifs in contact with RNA. Dotted lines show molecular interactions between protein and RNA.

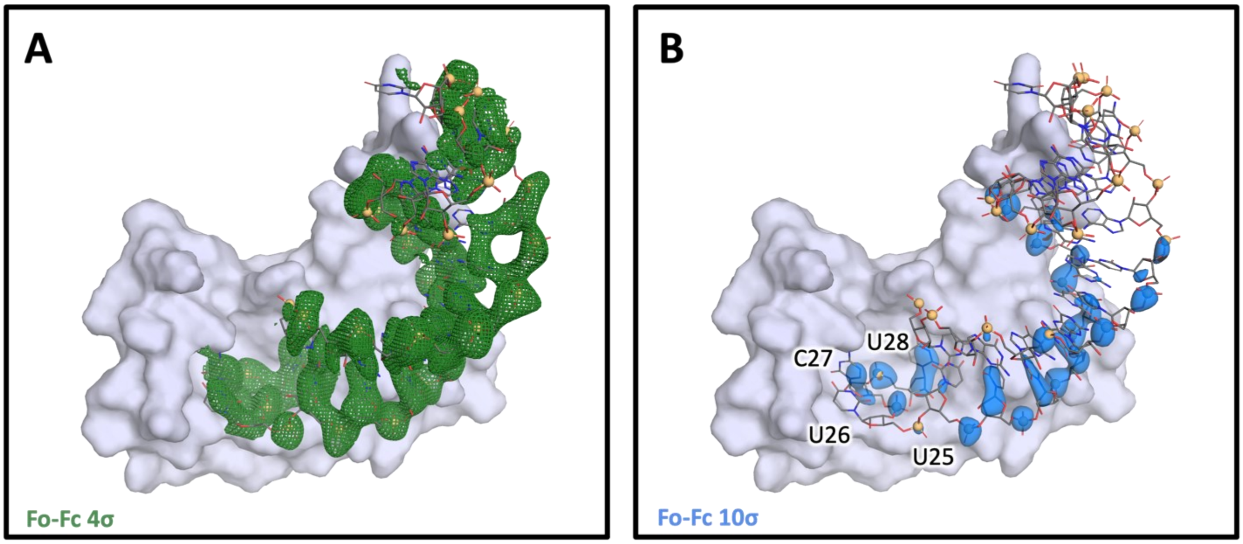


Figure 4-9. Difference electron density used to build the RocC$_{14\text{-}126}$:RocR$_{9bp\text{-}tet}$ complex.

The RocC/RocR crystal structure was initially solved by molecular replacement using the high resolution structure of RocC$_{14\text{-}126}$. Displayed is 4-fold averaged Fo-Fc density at 3.2 Å phased with the 10 RocC protomers in the asymmetric unit at 4 σ cutoff (green mesh, panel A) or 10 σ cutoff (blue surface, panel B) with the protein model displayed as a grey surface. The higher σ cutoff was used to help resolve the higher density phosphate groups from the bases and sugars in the RNA. The 4 nucleotides in the 3' tail are labeled in (B).

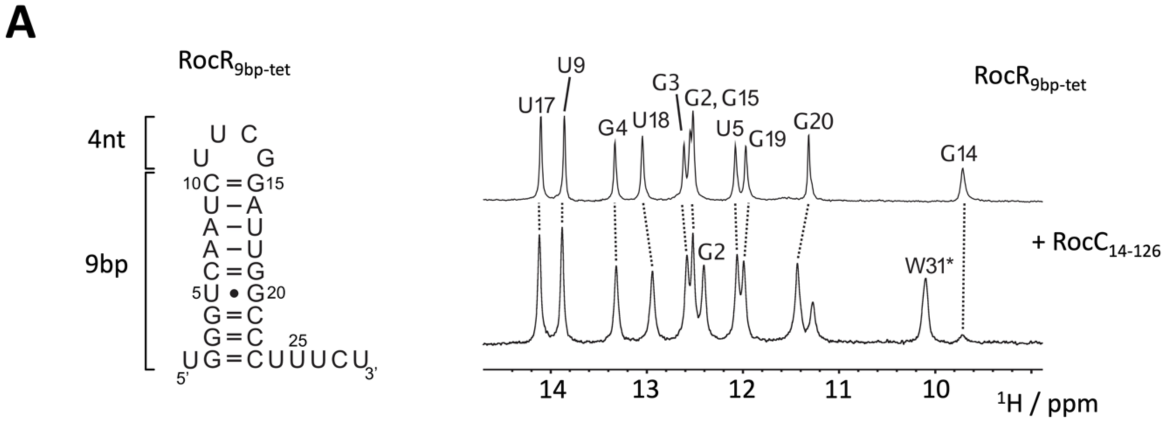


Figure 4-10. ^{1}H NMR spectra of the imino region of RocR$_{9bp\text{-}tet}$ and the RocR$_{SL3}$:RocC$_{14\text{-}126}$ complex.

The imino spectra indicate that the base-pairing pattern is the same for the free RNA and the complex. The unidentified signal at 11.3 ppm could not be connected to any imino residue but is believed to be a result of a specific non-Watson-Crick interaction either between protein and RNA or within the RNA, present only in the RNA:protein complex. This work was done by the collaboration with Kreutz and Tollinger laboratory, University of Innsbruck.

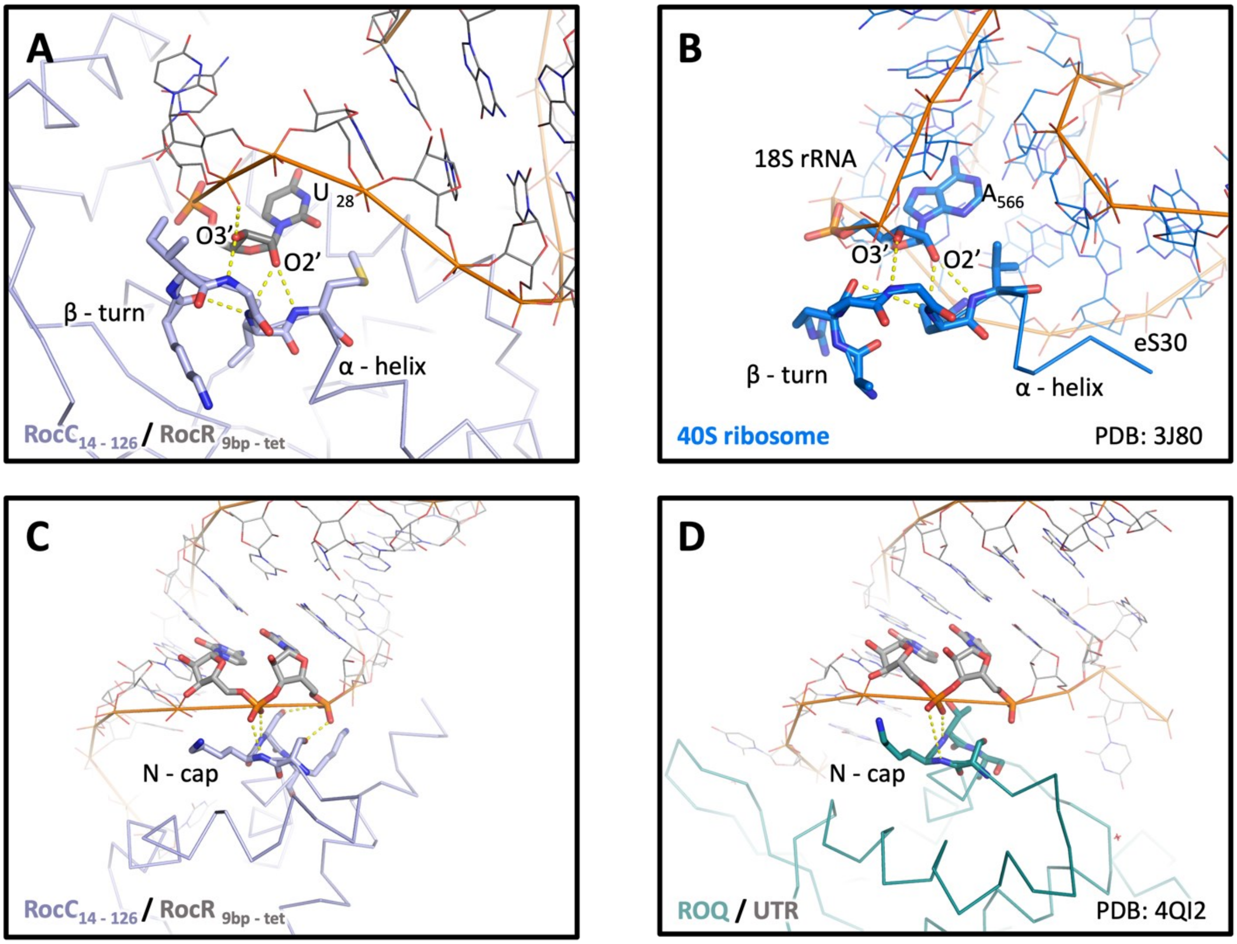


Figure 4-11. RNA recognition motifs in RocC are found in other RNA binding proteins.

(A) The N-cap motif of RocC recognizes phosphate groups along a single strand of the RocR hairpin stem. (B) A similar N-cap motif in ROQ recognizes the RNA phosphate backbone in a similar manner. Hydrogen bonds between the N-cap and the RNA phosphates are indicated with dashed bonds. Key residues involved in the interactions are shown as sticks. (C) The β-turn-α-helix motif in RocC recognizes the 3’ terminal nucleotide of RocR. (D) A similar β-turn-α-helix motif is found in the 40S ribosome between ribosomal protein eS30 and the 18S rRNA. In both cases, the O2’ and O3’ atoms of the C3’-endo ribose are hydrogen bonded to the motif in similar ways.

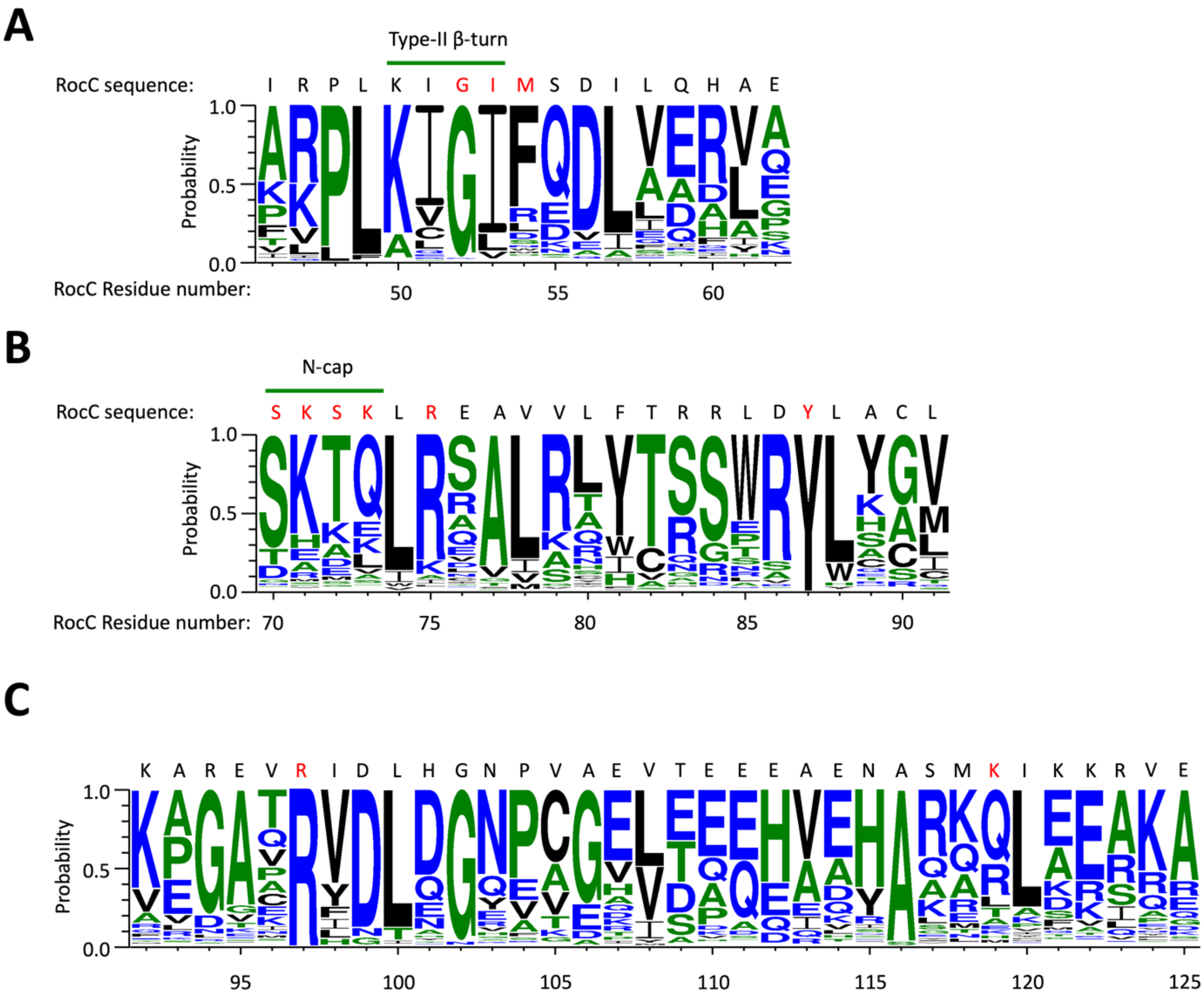


Figure 4-12. Graphical representation of sequence conservation in 674 FinO domain-containing proteins using WebLogo3 .

Red letters indicate the residues in contact with RocR$_{9bp\text{-}tet}$. (A) Sequence logo near the β-turn-α-helix motif. (B) Sequence logo near the N-cap motif. (C) Sequence logo for the C-terminal region of the RocC ProQ/FinO domain.

is nearly completely conserved in the ProQ/FinO domain family. This interaction effectively buries the 3' end of the strand in the protein and provides a mechanism for the selective recognition of a 3' terminal ribose sugar. We searched the protein-RNA structure database to find other examples of ribose recognition by similar β-turn-α-helix motifs (Chapter 2). The most similar example we found was from the 40S subunit of the eukaryotic ribosome where a β-turn-α-helix motif contacts a nucleotide within the 18S rRNA in the same orientation with the same hydrogen bonding arrangement (Figure 4-11A, B).

The hairpin portion of the RNA is bound by the N-terminal portion of α5, as well as supplementary interactions from α2 (Figure 4-8A, B). The N-terminus of α5 is capped by the highly conserved Ser70. The N-terminus of this helix hydrogen bonds to all 4 non-esterified oxygens in two successive phosphate groups – C_{21} and C_{22}. C_{21} phosphate is hydrogen bonded by Ser70 and Ser72, while the C_{22} phosphate is hydrogen bonded by the main chain NHs of Lys71 and Ser72. Both non-esterified oxygens of the C_{23} phosphate at the base of the stem are recognized by the side chains of Lys71 and Arg75. The G20 phosphate is recognized by Lys73, as well as by Ser21 from α2 and residues in more N-terminal regions of α2 may make further long-range electrostatic contacts with the backbone of the RNA 5' to this residue. Ser70, Lys71, Ser72 and Arg75 are all well-conserved in the ProQ/FinO domain, while Lys73 and Ser21 are less well-conserved. The precise recognition of all non-esterified oxygens in three consecutive phosphates along the same RNA strand presents an interesting possible mechanism for the recognition of an RNA duplex without direct interactions with both strands. We searched the protein:RNA and protein:DNA structure databases to uncover other examples of N-capped α-helices that recognize consecutive phosphates along a nucleic acid chain. We found that a very similar mechanism of RNA recognition is used by the ROQ domain of the mammalian Roquin protein (Schlundt et al., 2014; Tan et al., 2014), which binds 3 consecutive phosphates along a single strand of a hairpin RNA using a similar N-cap motif (Figure 4-11C, D). While many DNA binding proteins recognize DNA phosphates via hydrogen bonding interactions with α-helical N-terminal amide groups, we were unable to find any in which consecutive phosphates are recognized in a manner similar to either the ProQ/FinO or ROQ domains.

4.3.8. RocC:RocR crystal structure explains the conformation in solution

SAXS (small angle X-ray scattering) is a solution scattering technique that in theory can give structural information for macromolecular complexes (Putnam et al., 2007). We used SEC-SAXS, in which samples are first separated by size exclusion chromatography (SEC) before SAXS analysis. We used the SEC-SAXS beamline at ALS in which the SEC is installed in-line on the beamline. We prepared samples of $RocC_{1-126}$:$RocR_{SL3}$ and $RocC_{14-126}$:$SL3_{9bp-tet}$ for analysis. Both

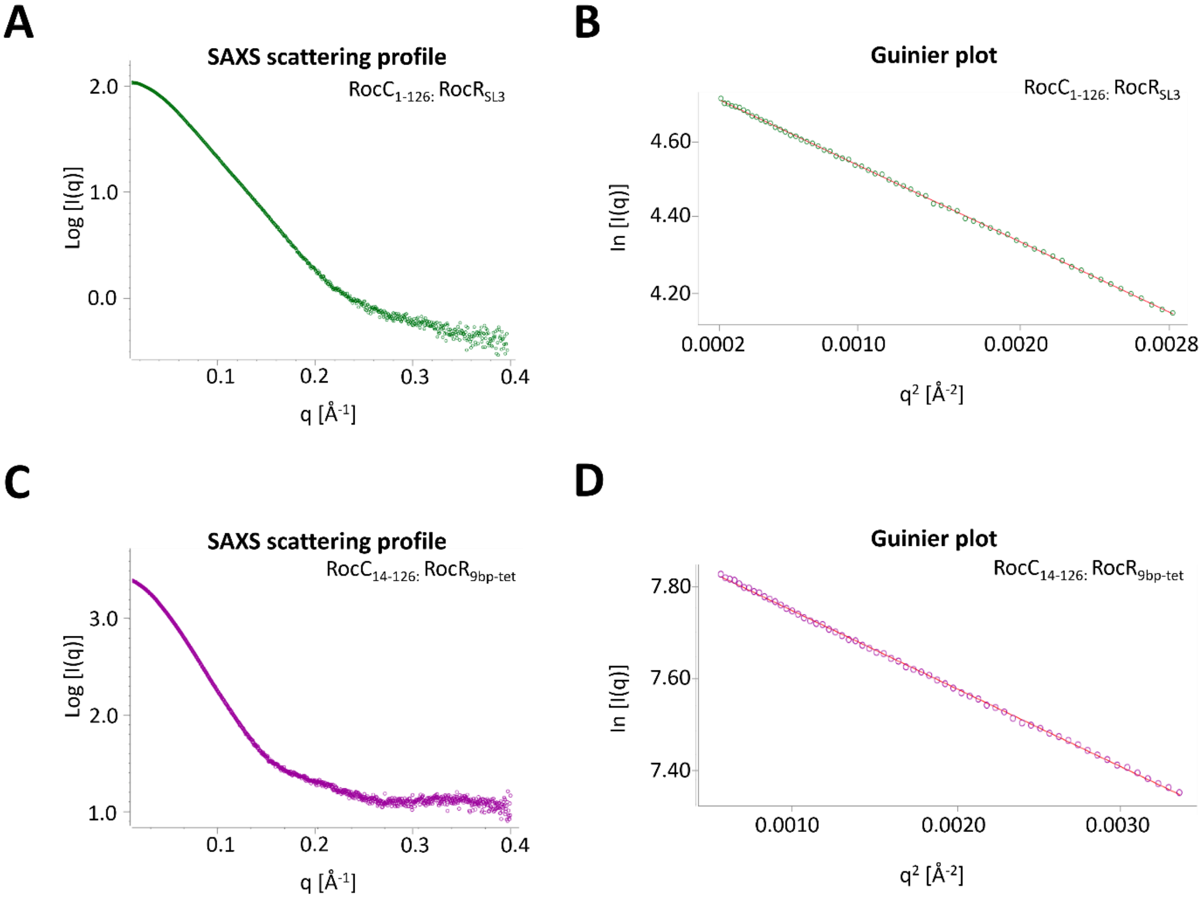


Figure 4-13. SAXS scattering profile and Guinier plot.

(A) SAXS scattering profile and (B) Guinier plot of $RocC_{1-126}$:$RocR_{SL3}$. (C) SAXS scattering profile and (D) Guinier plot of $RocC_{14-126}$:$RocR_{9bp-tet}$. Quality of Guinier plot is important for R_g (radius of gyration) calculation.

samples separated well on SEC and each gave scattering profiles that indicated little to no aggregation in the sample by Guinier analysis (Figure 4-13). To assess whether the scattering pattern was consistent with the RocC:RocR crystal structure, we calculated theoretical scattering curves (using MultiFoXS (Schneidman-Duhovny et al., 2016)) for the crystal structure and then calculated the goodness of fit of the theoretical and experimental curves. This was relatively straight-forward for the $RocC_{14-126}$:$SL3_{9bp-tet}$ complex. In this case, the theoretical curve calculated from the crystal structure fit the experimental data with a reasonable $\chi^2 = 2.65$, indicating that the crystal structure does approximate the solution structure. The $RocC_{1-126}$:$RocR_{SL3}$ complex, on the other hand, was more complicated. Singular value decomposition (SVD) can analyze the scattering

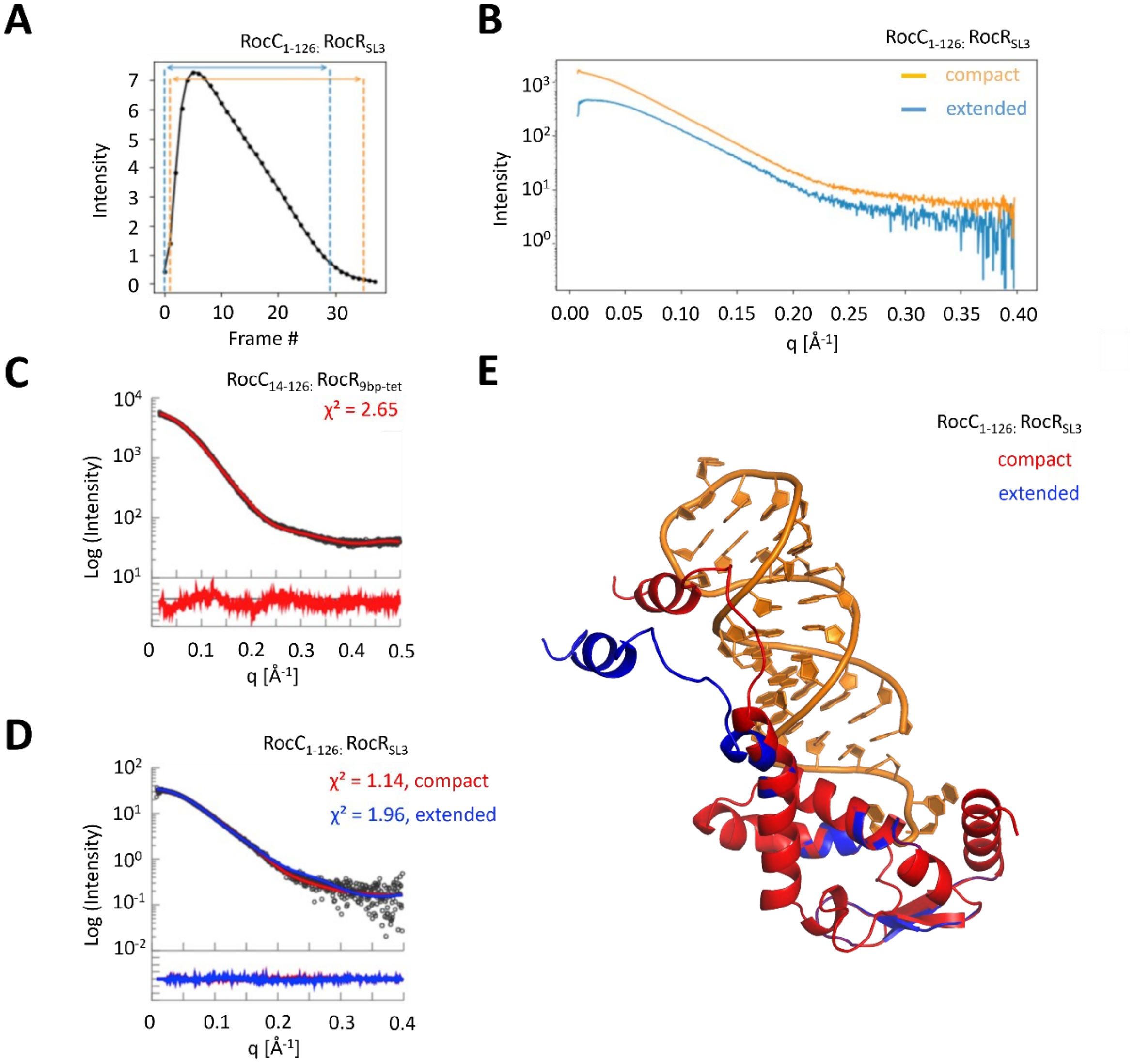


Figure 4-14. SEC-SAXS analysis of the FinO domain of RocC in complex with an intrinsic terminator.

(A) The size exclusion chromatogram of the $\mathrm{RocC}_{1\text{-}126}$:$\mathrm{RocR}_{\mathrm{SL3}}$ complex that was used in the SEC-SAXS analysis. The ranges of the extended (blue) and compact (yellow) species that were derived from the deconvolution process are indicated. (B) The deconvoluted scattering curves of the two species within the $\mathrm{RocC}_{1\text{-}126}$:$\mathrm{RocR}_{\mathrm{SL3}}$ complex are shown. (C) Fitting of the experimental (black) and calculated (red) scattering for the $\mathrm{RocC}_{14\text{-}126}$:$\mathrm{RocR}_{\mathrm{9bp\text{-}tet}}$ complex (D) Fitting of the experimental (black) and calculated (blue – extended; red - compact) scattering for the $\mathrm{RocC}_{1\text{-}126}$:$\mathrm{RocR}_{\mathrm{SL3}}$ complex. (E) 3D model of the extended (blue) and compact (red) $\mathrm{RocC}_{1\text{-}126}$:$\mathrm{RocR}_{\mathrm{SL3}}$ structures.

from different populations and evolving factor analysis (EFA) can deconvolute individual scatterings from an overlapped SEC-SAXS peak (Hopkins et al., 2017) (Figure 4-14A, B). We modeled these complexes using MultiFoXS, where we built the extended RNA stem and loop and also added a flexible region corresponding to RocC region 1-14. In MultiFoXS, the N-terminal region was allowed to move to maximize the fit of the model to the experimental data. In this way, we generated two models: one for the compact form ($\chi^2 = 1.14$, red curve, Figure 4-14D) and one for the more extended structure ($\chi^2 = 1.96$, blue curve, Figure 4-14D). The difference between the two models is that in the compact form, $\alpha 1$ is packed against the RNA hairpin, whereas in the more extended form, this helix extends away from the hairpin. In conclusion, the strong agreement of the experimental scattering curves with data calculated from the crystal structure suggested our crystal structure represent the conformation in solution.

4.4. Discussion

4.4.1. Optimization of crystallization conditions for RocC$_{14\text{-}126}$ in complex with RocR$_{9bp\text{-}tet}$

Crystallization is a trial and error process. Systematic approaches increase the chances of crystallization, as described in the introduction of this chapter, and these guided us in the crystallization of the apo-RocC and the RocC:RocR complexes.

Key factors for RocC$_{24\text{-}126}$ crystallization were low salt and high protein concentration. As discussed in the Introduction, salt is essential for the protein to be soluble in the buffer solution. However, high salt concentration increases the protein solubility and if it is too high, it causes "salting out" protein from the buffer, so it is essential to find optimal salt concentration between solubility and crystallization (McPherson & Gavira, 2014). DLS showed that RocC is monodisperse in salt concentrations from 30 – 150 mM, and indeed we were able to concentrate the protein to a high level under these conditions. The crystallization of NMB1681 at 91 mg/mL gave us a hint to test solubility, and we found RocC also does not precipitate up to 110 mg/mL. The insights gained from our preliminary data set the stage for its crystallization.

Two main factors that contributed to improving the quality of RocC$_{1\text{-}126}$ crystals were temperature and additives. Initial crystals grew at 16°C in 2 days; however, crystals did not form over 3 months at 4°C. 4°C potentially gives long-term stability of RocC$_{1\text{-}126}$ and slows down the evaporation rate in the droplet. The role of additive screens was unclear compared to temperature because none of additives were visualized in the crystal structure. The nine copies of $\text{RocC}_{1\text{-}126}$ were stabilized by crystal contacts with ammonium ions, which were in the reservoir solution. Furthermore, the same crystal habit was grown with different additives, such as multivalent ions, polymer, carbohydrate, and organic solvents (Table 4-1B). Nevertheless, additives did ultimately produce crystals that diffracted at high resolution.

We determined the structure of $\text{RocC}_{14\text{-}126}$:$\text{RocR}_{9\text{bp-tet}}$ at a relatively modest resolution of 3.2 Å. Our trials produced four different crystals from two complexes: three from $\text{RocC}_{14\text{-}126}$:$\text{RocR}_{9\text{bp-tet}}$ and one from $\text{RocC}_{1\text{-}126}$:$\text{RocR}_{\text{SL3}}$. The main issue was that our two reproducible crystals had poor diffraction, and we failed to reproduce two crystals with better diffraction than in the initial screening. We managed to determine the structure from non-reproducible crystals. Intriguingly, the $\text{RocC}_{14\text{-}126}$:$\text{RocR}_{9\text{bp-tet}}$ complex we crystallized had only 4 copies bound to RNA and another 6 copies were apo-RocC. Even though the SEC-MALS study showed the formation of stable one-to-one complex, 60% of protomers in the lattice were not bound to RNA. 40% RNA occupancy suggested the possibility there may be a heterogeneous population in the crystallization drop. The crystal structure showed that only 11 residues on the 3' side of RocR_{SL3} had contact with RocC. The 5' side of RocR_{SL3} is near the saddle region of RocC (helix 5), which is highly positively charged. If our RocR_{SL3} constructs had a longer 5' extension or included part of RocR_{SL2}, this extended 5' region might possibly interact with residues Arg83 and Arg84 from helix 5. The additional interaction of the extended 5' side with protein could give stronger binding, higher homogeneity and it could enhance the crystallization. However, our efforts to measure the binding of intact RocR to RocC indicate that addition of the 5’ regions do not measureably enhance binding affinity (Attaiech et al., 2016). The 5’ region is moreover critical for interactions with target mRNAs such as *comEA*, so further studies with more extended RNAs might yield insights into how RocC facilitates RNA:RNA association.

4.4.2. RocC:RocR provides a model for the recognition of intrinsic transcriptional terminators by proteins with a ProQ/FinO domain

The structure of the RocC:RocR_{SL3} complex presented here reveals a mechanism where RocC specifically binds its natural partner RocR by interacting with its terminator, i.e. the hairpin of SL3 and its 3' single-stranded polypyrimidine tail. This structure also provides a model to understand RNA recognition by other ProQ/FinO domain proteins. The pocket that binds the 3' nucleotide of the terminator is particularly well conserved in the ProQ/FinO family, not only at the sequence level, but it is also conformationally conserved in also FinO and NMB1681. The position of the N-cap motif is absolutely conserved in the ProQ/FinO family, and Arg75 is almost completely conserved (80% identity) with the major substitution being lysine (Figure 4-12B). Furthermore, the good agreement of the crystal structure of $\mathrm{RocC_{14\text{-}126}}$:$\mathrm{RocR_{9bp\text{-}tet}}$ complex with solution SEC-SAXS data (Figure 4-13), indicates that the static FinO:intrinsic terminator structure is relevant in solution.

The structure explains previous biochemical and mutagenesis data on FinO as well as ProQ. Site-specific protein-RNA crosslinking identified residues in and around the conserved 3' nucleotide binding pocket (Lys125, Lys165), as well as the N-cap (Arg121) as key contact points for FinO (Ghetu et al., 2002). Likewise, bacterial 3-hybrid experiments identified the same surfaces on ProQ as important to recognize transcriptional terminator structures (Pandey et al., 2020). Ribonuclease footprinting revealed dramatic protection of the FinP 3' tail by FinO, as well as protection of the first 3-4 nucleotides within the stem, precisely the same region that is bound by RocC in the RocC:RocR structure (Arthur et al., 2011). Interestingly, it was previously shown that the recognition of FinP by FinO is strongly dependent on the chemical structure of the 3' nucleotide (Figure 4-15) (Arthur et al., 2011). The recognition of the 3' hydroxyl group by RocC should be tested.

The ProQ/FinO family members that have been studied in detail (FinO, RocC, ProQ and NMB1681) all bind transcriptional terminator structures but with varying degrees of specificity. While FinO and RocC bind to just one or two physiological partners, ProQ and NMB1681 bind to dozens of different sRNAs. Our structure of the RocC:$\mathrm{RocR_{SL3}}$ complex suggests a mechanism for the specific recognition of the hairpin as well as for the 3' polypyrimidine tail. The duplex portion of the hairpin is recognized on one strand by the N-cap motif. Specificity for the duplex is likely conferred through recognition of the A-form geometry of the contiguous phosphate groups, rather than through a direct recognition of the 5' strand of the hairpin. The 3' polypyrimidine strand adopts a hook-like structure, stabilized by hydrogen bonding between the penultimate phosphate and the terminal 3’-hydroxyl (Figure 4-8A, C). In this conformation, the strand tracks into the 3'

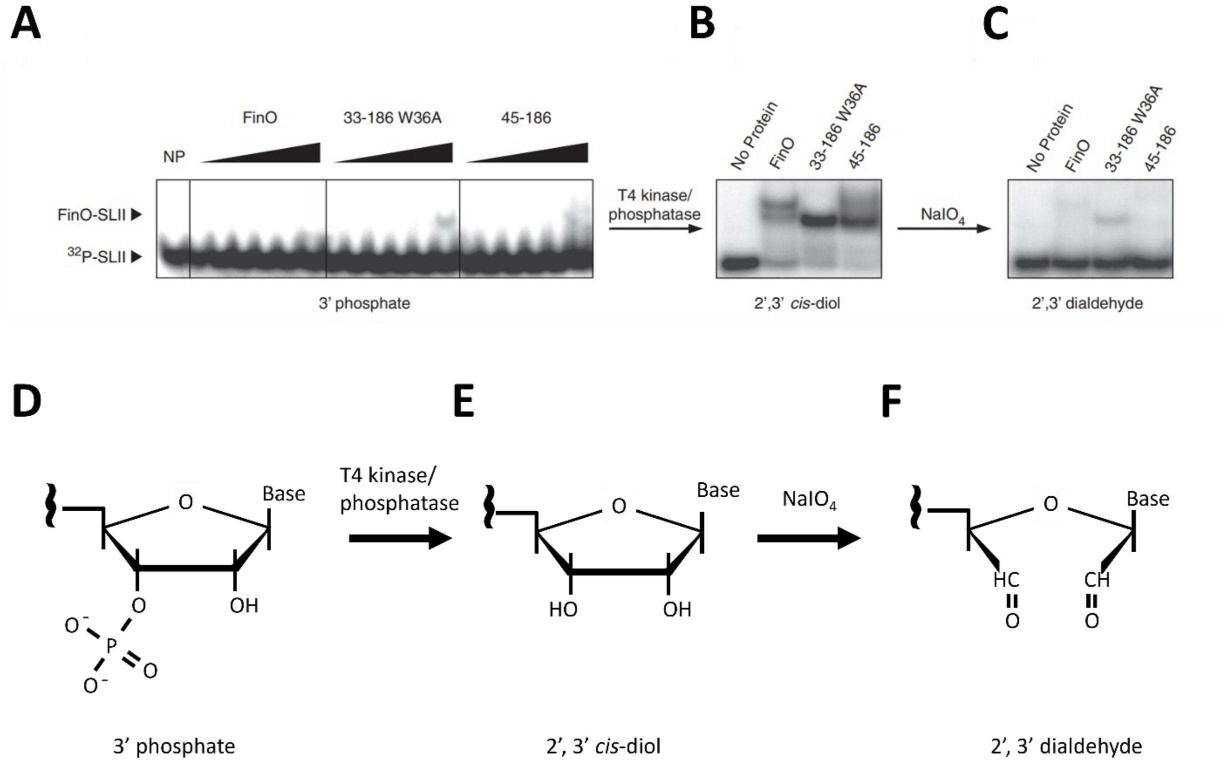


Figure 4-15. 3' hydroxyl group recognition by FinO protein.

(A) Gel shift assay of 3' phosphorylated SLII in titration with gradient of FinO, FinO$_{33\text{-}186}$ W36A, and FinO$_{45\text{-}186}$. (B) Gel shift assay of 2', 3' *cis*-diol SLII. (C) Gel shift assay of 2', 3' dialdehyde SLII. (D) Chemical structure of 3' phosphate, (E) 2', 3' *cis*-diol, (F) 2', 3' dialdehyde. Figure (A-C) was taken from previous publication (Arthur et al., 2011).

nucleotide binding pocket. While it is difficult to definitively model side chain – base hydrogen bonding interactions at this resolution, the structure does suggest some possibilities that may indicate specificity for pyrimidines within the tail. Highly conserved Arg75 participates in RNA backbone recognition, but may also hydrogen bond with the 5'-most uridines of the polypyrimidine tail. The next 2 nucleotides are packed against $\alpha 7$ in a manner that may sterically favour pyrimidines over purines, and residues such as Lys119 and Asn115 may provide sequence specificity through hydrogen bonding interactions with these bases. Consistent with this idea, mutations at these residues reduce binding affinity and transformation efficiency. The uridine base of the terminal 3' nucleotide packs against Thr82 and mutation of this residue does significantly reduce transformation efficiency.

The structural determination of RocC:RocR complex gave an insight into molecular mechanism how the FinO domain recognizes its binding partner. The observed interactions in the structural analysis are confirmed by site-directed mutagenesis *in vitro* and *in vivo* described in the next chapter. In addition, the structure raised new questions:

1. Is a 10 nucleotide single-stranded RNA sufficient to interact with FinO domain since only these residues are directly contacted by the protein?

2. Does the 3' hydroxyl group play a key role in RocC:RocR interaction?

3. Do these proteins have a preference for the length of the single-stranded 3’ tail?

These issues are addressed in the next chapter.

Chapter 5.

Mechanistic insights into how ProQ/FinO domains bind RNA transcriptional terminators

5.1. Overview

ProQ/FinO domains selectively bind RNA transcriptional terminators composed of a hairpin linked to a 3’ single stranded tail. Our crystal structure of the RocC ProQ/FinO domain bound to the RocR terminator structure showed that the 3’ nucleotide of the terminator is bound by a conserved pocket in the domain, and the hairpin portion is bound by a N-capped helical motif. Here we show that point mutations within these two binding regions impact binding affinity and reduce the ability of RocC to repress the uptake of extracellular DNA in *Legionella*. Our structure showed that RocC only contacts a single strand of the hairpin, raising the question as to why these proteins specifically bind hairpin RNAs and not ssRNAs. We suggest that the binding preference for hairpin containing binding partners is due to recognition of the A-form conformation of one strand of the hairpin by the N-cap motif and our thermodynamic analysis of the binding interaction indicates that this is driven by the entropic savings from hairpin binding over more flexible ssRNA partners. We furthermore show that the ProQ/FinO domains of FinO and ProQ can also bind RocR-derived RNA targets and that ProQ, like RocC, binds preferentially to targets with a 5 nucleotide 3’ single-stranded tail. Finally, we show that phosphorylation of the 3’ nucleotide abrogates interactions of RocC with the RocR terminator, underlining the critical importance of the structure of the 3’-nucleotide and its recognition by the β-turn-α-helix motif.

5.2. Introduction

The ProQ/FinO family of RNA chaperones regulate post-transcriptional gene expression throughout proteobacteria. Intriguingly, different ProQ/FinO proteins seem to have quite different levels of biological specificity. While certain family members bind dozens of RNA targets within cells, others are highly specific and only bind a single target. The plasmid-encoded FinO protein in *E. coli* specifically binds to FinP, the antisense RNA of the *traJ* mRNA 5’ UTR, to regulate TraJ protein expression. TraJ is the positive transcriptional activator of plasmid genes that promote bacterial conjugation. FinO binds FinP through a structured stem loop (Arthur et al., 2011) and protects FinP from RNase degradation (Jerome et al., 1999). In addition, FinO promotes the initiation of loop-loop kissing interactions between FinP and *traJ* mRNA 5’ UTR (Arthur et al., 2003). Duplex formation of the sense-antisense RNA pair inhibits translation by blocking the ribosomal binding site in *traJ* mRNA, and the inhibition of *traJ* gene translation downregulates conjugation. The chromosomally encoded RocC in *L. pneumophila* is another example of a specific binder, and it works *in trans*. Unlike the perfect complementarity between sense and antisense RNAs, the imperfect base-pairing of the RocR *in trans* system enabled targeting four competence related genes: *comEA, comEC, comF, and comM* (Attaiech et al., 2016). These four mRNAs have a consensus "RocR box" sequence on on a predicted exposed loop, which can initiate limited base-pairing with the target mRNA 5' UTRs. RocC plays a role in protecting RocR from degradation and promoting RocR-target RNA hybridization.

On the other hand, other ProQ/FinO proteins act as global RNA binding proteins, such as ProQ in *S. enterica*, *E. coli*, and *N. meningitidis* (NMB1681) (Bauriedl et al., 2020; Holmqvist et al., 2018; Smirnov et al., 2016). UV-CLIP combined with RNA sequencing revealed that these global RNA binders each target multiple structured sRNAs, and mainly mRNA 3' UTRs (Bauriedl et al., 2020; Holmqvist et al., 2018). Rifampicin stability assays showed the half-life of the targeted RNAs decrease in the absence of the ProQ protein in *S. enterica* and *N. meningitidis*. Further experiments showed that ProQ interaction with mRNA 3' UTR protects RNA from 3'-5' exoribonucleolytic degradation *in vivo* and *in vitro*.

The conserved ProQ/FinO domain has a concave face that is often positively charged and has been speculated to be the main RNA binding interface (Arthur et al., 2011; Ghetu et al., 2002; Olejniczak & Storz, 2017). The determination of the FinO structure (Ghetu et al., 2000) allowed investigating residues in proximity to FinP using crosslinking analysis (Ghetu et al., 2002). The most efficiently crosslinked residues were Arg121, Lys125, and Arg165, which are on the concave surface of the FinO. Random mutagenesis approaches in *E. coli* ProQ mapped residues that impaired binding *in vivo.* Bacterial 3-hybrid (B3H) revealed that most of the point mutations, which are defective in binding, are on the concave surface of the ProQ in *E. coli* (Pandey et al., 2020). Biochemical and genetic approaches expanded the understanding of the important residues on the concave surface for the binding in ProQ/FinO domain. However, the molecular recognition of the protein:RNA remained unclear due to the lack of complex structure.

In this chapter, we generated site-directed mutagenesis on the RocC residues based on structural studies for *in vitro* binding and *in vivo* functional studies. The influences of mutations in binding were tested with 5' FAM labeled RocR$_{SL3}$ using fluorescence polarization (FP). The mutations of on the N-cap motif and the concave pocket surrounding 3' tail showed the critical residues for RNA interaction, as well as for the repression of natural transformation *in vivo*. We also tested whether ssRNA is sufficient to interact with ProQ/FinO family proteins using FinO, ProQ, and RocC. FinO showed similar binding affinity to both the RocR intrinsic terminator and ssRNA, however, ProQ showed ~10-fold tighter binding to the intrinsic terminator. RocC showed a dramatic decrease in binding to ssRNA compared to the intrinsic terminator, $\mathrm{RocR}_{\mathrm{9bp\text{-}tet}}$. Furthermore, we demonstrated the critical importance of the presence of the free 3'-OH on the terminal residue as no binding was observed to $\mathrm{RocR}_{\mathrm{9bp\text{-}tet}}$ with a 3' phosphate group. We also tested the tail length preferences of RocC, FinO and ProQ.

5.3. Results

5.3.1. Site-directed mutagenesis reveals the importance of specific RocC:RocR contacts for binding and DNA uptake *in vivo*

Guided by the RocC:RNA structure (Figure 4-8E and Figure 5-1A), we created a set of RocC mutants to test the contribution of individual amino acid – RNA contacts to the RocC:RocR interaction. *In vitro*, we assessed the binding affinity between $\mathrm{RocR}_{\mathrm{9bp\text{-}tet}}$ and each $\mathrm{RocC}_{14\text{-}126}$ mutant via a FP assay (Figure 5-1B, E and Table 5-1A). *In vivo*, the interaction between RocC and RocR is essential to stabilize RocR and promote the post-transcriptional repression of genes required for the uptake of DNA from the environment (transformation). Consequently, a strain in which RocC is absent or non-functional is more transformable than its WT counterpart ("hypercompetent" phenotype) (Attaiech et al., 2016). We thus monitored the effects of different point mutations of RocC by testing the transformability of *L. pneumophila* strains expressing these variants compared to the WT protein (Figure 5-2A).

Mutation of residues within and surrounding the 3' nucleotide binding pocket yielded a significant impact on the RocC:RocR binding affinity and transformation efficiency. Mutation of the highly conserved Arg97 (R97M), which forms the base of the 3' nucleotide binding pocket, resulted in an ~200-fold reduction in binding affinity as well as a dramatic reduction in transformation repression similar to that observed in the ΔRocC or ΔRocR controls (Figure 5-1B, 5-2A). We mutated the absolutely conserved Tyr87 to phenylalanine (Y87F). *In vivo*, this mutation resulted in a complete loss of repression of transformation; *in vitro* this mutant was

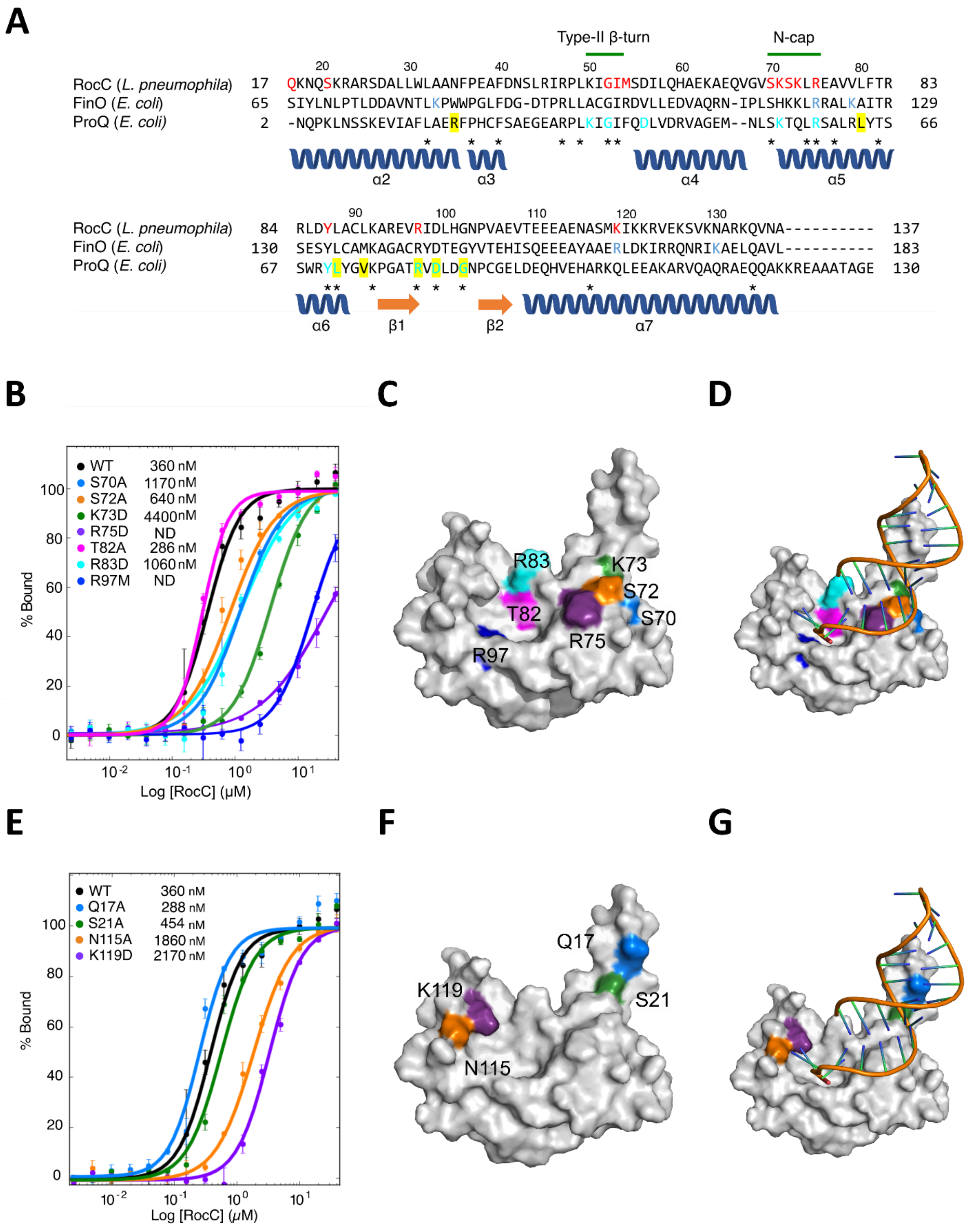

A
Type-II β-turn
N-cap
20 30 40 50 60 70 80
RocC (L. pneumophila) 17 QKNQSKRARSDALLWLAANFPEAFDNSLRIRPLKIGIMSDILQHAEKAEQVGVSKSKLREAVVLFTR 83
FinO (E. coli) 65 SIYLNLPTLDDAVNTLKPWWPGLFDG-DTPRLLACGIRDVLLEDVAQRN-IPLSHKKLRRALKAITR 129
ProQ (E. coli) 2 -NQPKLNSSKEVIAFLAERFPHCFSAEGEARPLKIGIFQDLVDRVAGEM--NLSKTQLRSALRLYTS 66
α2
α3
α4
α5
90 100 110 120 130
RocC (L. pneumophila) 84 RLDYLACLKAREVRIDLHGNPVAEVTEEEAENASMKIKKRVEKSVKNARKQVNA---------- 137
FinO (E. coli) 130 SESYLCAMKAGACRYDTEGYVTEHISQEEEAYAAERLDKIRRQNRIKAELQAVL---------- 183
ProQ (E. coli) 67 SWRYLYGVKPGATRVDLDGNPCGELDEQHVEHARKQLEEAKARVQAQRAEQQAKKREAAATAGE 130
α6
β1
β2
α7
B
WT 360 nM
S70A 1170 nM
S72A 640 nM
K73D 4400 nM
R75D ND
T82A 286 nM
R83D 1060 nM
R97M ND
% Bound
0
20
40
60
80
100
10^-2
10^-1
10^0
10^1
Log [RocC] (µM)
C
R83
K73
S72
T82
S70
R75
R97
D
E
WT 360 nM
Q17A 288 nM
S21A 454 nM
N115A 1860 nM
K119D 2170 nM
% Bound
0
20
40
60
80
100
10^-2
10^-1
10^0
10^1
Log [RocC] (µM)
F
Q17
K119
S21
N115
G

Figure 5-1. Effects of RocC point mutations on RocR binding using FP.

(A) Structure-based sequence alignment of the ProQ/FinO domains of RocC, FinO, and ProQ. Red letters in RocC indicate residues in contact with $\mathrm{RocR}_{9\mathrm{bp\text{-}tet}}$ in the crystal structure. Asterisk (*) are highly conserved residues across the three proteins. Blue letters in the FinO sequence indicate residues that show strong cross-linking with SLII of FinP (Ghetu et al., 2002). Cyan letters in the ProQ sequence show residues that are critical for RNA binding in 3-hybrid screening (Pandey et al., 2020). Yellow highlights in ProQ indicate vital residues for ProQ function (Rizvanovic et al., 2021). The indicated secondary structure is derived from the RocC:RocR crystal structure. (B) FP binding assay for 5' FAM-labeled $\mathrm{RocR}_{\mathrm{SL3}}$ with different $\mathrm{RocC}_{14\text{-}126}$ point mutants on the concave surface. (C, D) Surface representation of $\mathrm{Roc}_{14\text{-}126}$ alone (C) and with RocR (D) colored to indicate the positions of mutated residues. (E) FP binding assay for 5' FAM-labeled $\mathrm{RocR}_{\mathrm{SL3}}$ with different $\mathrm{RocC}_{14\text{-}126}$ point mutants on the N-terminus and the C-terminus. (F, G) Surface representation of $\mathrm{Roc}_{14\text{-}126}$ alone (F) and with RocR (G).

insoluble, suggesting a significant folding defect related to the mutation of this buried residue. Mutation of residues surrounding the 3' nucleotide binding pocket showed more subtle effects. Mutations of residues on α7, N115A and K119D, showed less dramatic but still significant reductions (~5-fold) in binding affinity, as well as significant reductions in transformation repression (Figure 5-1E, 5-2A). Asn115 is positioned within hydrogen bonding distance to the U_{26} and/or C_{27} base of SL3 and is most often a His or Tyr within the ProQ/FinO family (Figure 4-12C). Lys119 is positioned to make a cation-π interaction with the U_{26} base. However, within the ProQ/FinO family this residue is most often a Gln or Arg (Figure 4-12C). Arg83 also lines the 3' nucleotide binding pocket and is positioned to contact the 5'-most nucleotide. Mutation of this residue (R83D) led to a modest (~3-fold) reduction in binding affinity, as well as a reduction in transformation repression. Arg is commonly observed at this position in ProQ/FinO domains, however Ser is the most conserved residue at this position (Figure 4-12B). Mutation of two other residues in the pocket led to no significant reduction in binding affinity. Thr82 packs against the base of the terminal U_{28} base, however, mutation of this residue to an alanine did not reduce binding, even though Thr is highly conserved at this position in ProQ/FinO domains (Figure 4-12B). This mutation did lead to a significant reduction in transformation repression which indicates that this residue is important for RocC to function properly. Interestingly, mutation of the highly conserved Ile51(Figure 4-12A), which packs against the backbone of C_{27}, to an alanine did not result in any reduction in binding and had a very limited impact, if any, on transformation repression.

Mutation of residues in the N-cap RNA binding motif also led to significant reductions in binding and biological activity. The most dramatic effect was observed for R75D which led to a ~100-fold reduction in binding affinity, as well as a complete loss of the mutant protein's ability to repress transformation. K71D and K73D also displayed significant reductions in binding affinity (~10-fold) as well as an almost complete loss in transformation repression. S70A resulted in less pronounced but still significant defects in binding and transformation repression.

We also tested mutations in the N-terminal helix α2, which our structure suggested could make limited contacts to the RNA hairpin. Deletion of the N-terminal 13 residues, which are disordered in our structures, did not impact RNA binding affinity or transformation repression. However further deletion to residue 24 resulted in a ~3-fold reduction in binding affinity, similar to what we had observed by EMSA, and a total loss of transformation repression (Figure 5-2B). Individual mutation of direct contacted residues in structural analysis (Q17A and S21A) and positively charged residues (K15D and K18D) in the 14-24 region did not exhibit reduced RNA binding, nor did they cause significant defects in transformation repression *in vivo* (Figure 5-1E, 5-2A, Table 5-1A). Furthermore, we tested the binding affinity of a RocC construct containing an additional C-terminal predicted helical region ($\text{RocC}_{1\text{-}137}$). This construct had the same binding affinity as 1-126, indicating that this additional region does not play a significant role in RNA binding (Figure 5-3D, Table 5-1A).

5.3.2. ProQ/FinO domain of RocC recognizes the geometry of RNA and the terminal 3' hydroxyl group

Structural analysis showed that the ProQ/FinO domain of RocC contacts only 11 nucleotides on the 3' side of RocR_{SL3} and does not directly contact both strands of the hairpin (Figure 4-8D, E). Paradoxically however, our previous work on FinO showed that FinO bound to hairpin-containing RNAs and with only low affinity to ssRNAs (Arthur et al., 2003). We hypothesized that the specificity of ProQ/FinO domains for hairpin-containing RNAs might come from the recognition of the A-form conformation of the RNA, and not necessarily from a direct recognition of both strands *per se*. Our analysis suggested this might be possible because the N-cap motif coordinated multiple contiguous phosphates along the single strand that could only be

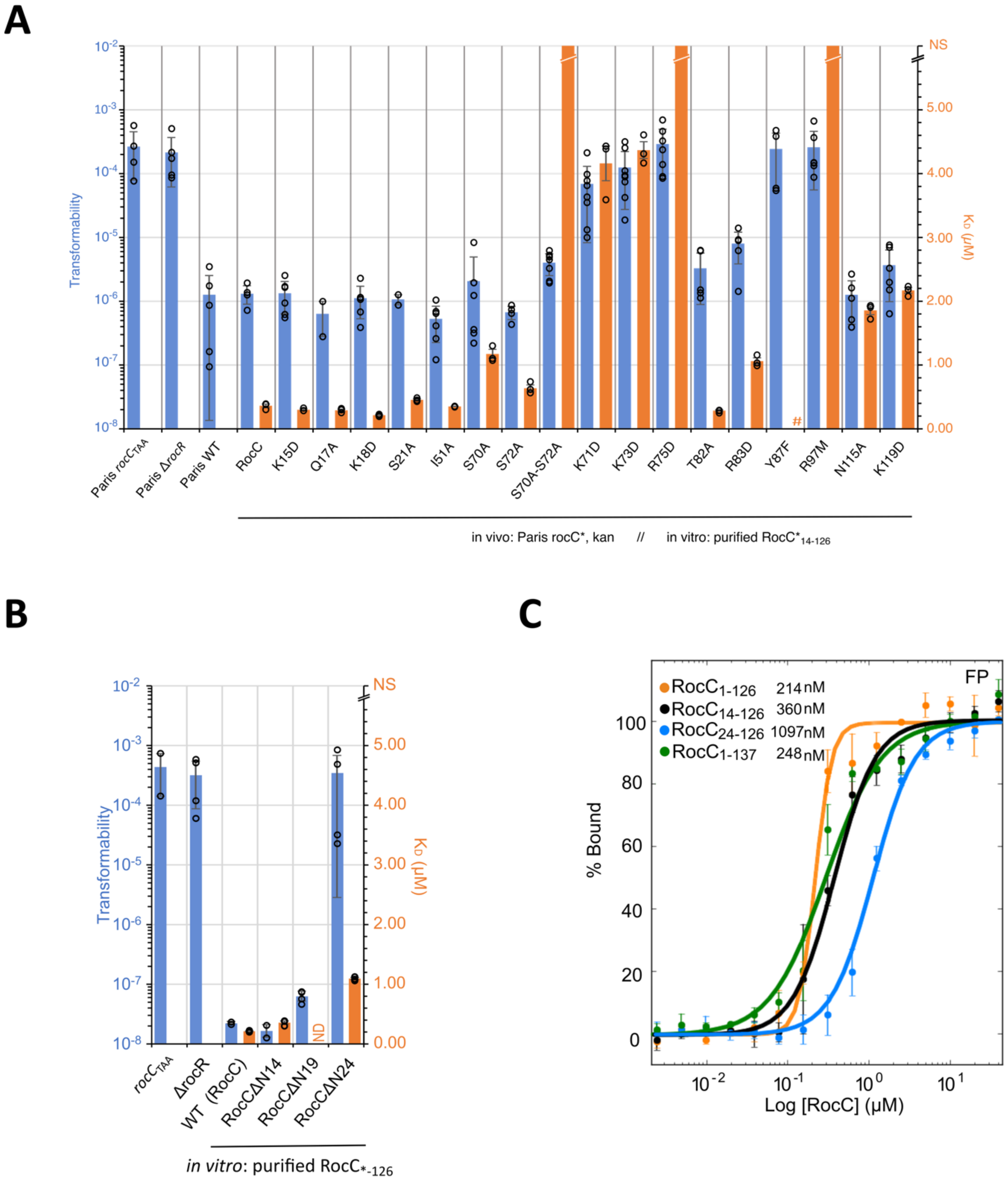

A
10-2
10-3
10-4
10-5
10-6
10-7
10-8
Transformability
NS
5.00
4.00
3.00
2.00
1.00
0.00
KD (µM)
#
Paris rocCTAA
Paris ΔrocR
Paris WT
RocC
K15D
Q17A
K18D
S21A
I51A
S70A
S72A
S70A-S72A
K71D
K73D
R75D
T82A
R83D
Y87F
R97M
N115A
K119D
in vivo: Paris rocC*, kan // in vitro: purified RocC*14-126
B
10-2
10-3
10-4
10-5
10-6
10-7
10-8
Transformability
NS
5.00
4.00
3.00
2.00
1.00
0.00
KD (µM)
ND
rocCTAA
ΔrocR
WT (RocC)
RocCΔN14
RocCΔN19
RocCΔN24
in vitro: purified RocC*-126
C
RocC1-126 214 nM
RocC14-126 360 nM
RocC24-126 1097 nM
RocC1-137 248 nM
FP
100
80
60
40
20
0
% Bound
10-2
10-1
100
101
Log [RocC] (µM)

Figure 5-2. Correlation of RocC:RocR binding *in vitro* and functional relevence *in vivo*.

(A) A graph displaying RNA binding affinities and transformation efficiencies for the indicated strains and RocC mutants. The orange columns indicate the K_D values measured by FP, and the blue columns indicate the relative transformability of the indicated mutant or strain. # indicates a mutant which could not be purified due to low protein solubility. NS indicates mutants where RNA binding could not be detected. Values for the individual transformation measurements are shown, and the standard deviations from three independent measurements are shown for the FP binding data. (B) Role of the N-terminus of RocC for $RocR_{SL3}$ binding in vitro and for uptake of DNA (transformation) *in vivo*. Transformation was assessed in *Legionella* strains containing either WT RocC, a mutant in which rocC translation is disrupted (*$rocC_{TAA}$*), a RocR deletion (ΔrocR), or different deletions at the N-terminus (RocCΔN14 - deletion of a.a. 1-13; RocCΔN19 – deletion of a.a. 1-18; RocCΔN24 – deletion of a.a. 1-23). The binding measurements were carried out with a fluorescence polarization (FP) assay using $RocC_{1-126}$ or the indicated N-terminal deletion mutants, with FAM-labelled $RocR_{SL3}$ as a target. (C) FP binding assays for 5' FAM-labeled $RocR_{SL3}$ with different RocC truncation mutants.

A

RocC	Mutant	K_D (μM)	Standard error (μM)	p-value
1-126	WT	0.214	0.007	0.1381
24-126	WT	1.097	0.019	<0.0001
1-137	WT	0.248	0.017	N/A
14-126	WT	0.36	0.03	0.0314
14-126	K15D	0.298	0.014	0.1344
14-126	Q17A	0.288	0.019	0.1125
14-126	K18D	0.213	0.009	0.0094
14-126	S21A	0.454	0.016	0.0506
14-126	I51A	0.3502	0.0032	0.7616
14-126	S70A	1.17	0.07	0.0004
14-126	K71D	4.2	0.3	0.0002
14-126	S72A	0.64	0.05	0.0086
14-126	K73D	4.4	0.1	<0.0001
14-126	R75D	ND	ND	ND
14-126	T82A	0.286	0.012	0.0838
14-126	R83D	1.06	0.05	0.0003
14-126	R97M	ND	ND	ND
14-126	N115A	1.86	0.07	<0.0001
14-126	K119D	2.17	0.05	<0.0001
14-126	S70/72A	ND	ND	ND

B

Protein	RNA Mutant	K_D (μM)	Standard error (μM)
$RocC_{14-126}$	DNA	ND	ND
$RocC_{14-126}$	$RocR_{3nt}$	8.30	0.27
$RocC_{14-126}$	$RocR_{5nt}$	0.374	0.029
$RocC_{14-126}$	$RocR_{8nt}$	14.4	1.8
$FinO_{45-186}$	$RocR_{3nt}$	2.02	0.11
$FinO_{45-186}$	$RocR_{5nt}$	0.903	0.078
$FinO_{45-186}$	$RocR_{8nt}$	1.03	0.82
$ProQ_{1-130}$	$RocR_{3nt}$	0.95	0.12
$ProQ_{1-130}$	$RocR_{5nt}$	0.289	0.018
$ProQ_{1-130}$	$RocR_{8nt}$	1.469	0.081

Table 5-1. Effects of RocC point mutations on RNA binding and isolated ProQ/FinO domain binding preferences in vitro using FP. (A) and (B) Tables of K_D values derived from FP binding assays. ND indicates a K_D could not be determined because binding saturation was not achieved. p-values indicated here compare K_D of the $RocC_{14-126}$ point mutation to the WT. We considered a $p\text{-value} < 0.05$ to be statistically significant.

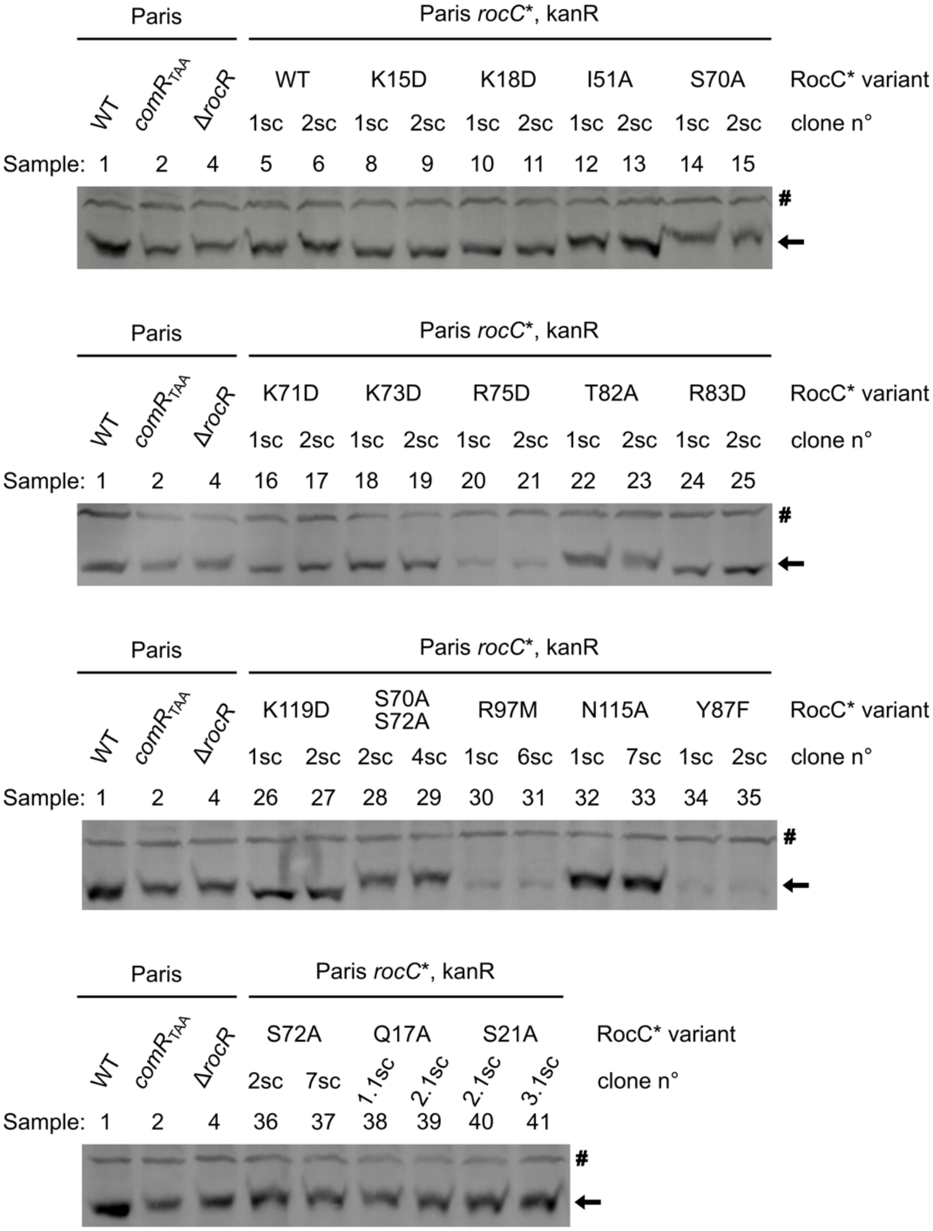


Figure 5-3. Detection of RocC* variants *in vivo*.

Western-blot analysis of RocC and RocC* variants in the Paris WT strain, the Δ*rocR* and *comR*$_{\mathrm{TAA}}$ hypercompetent strains and the Paris *rocC*, kanR strain and its derivatives bearing a mutated allele of *rocC* instead of the WT allele. The strains were grown up to OD_{600}~1 in AYE medium @ 37°C. Samples of $1.5.10^8$ cells were analysed by SDS-PAGE and Western-blot using polyclonal anti-RocC antibodies. The symbol # indicates a cross-reactive band that can be used as a loading control, the arrow points to the bands corresponding to the RocC* proteins.

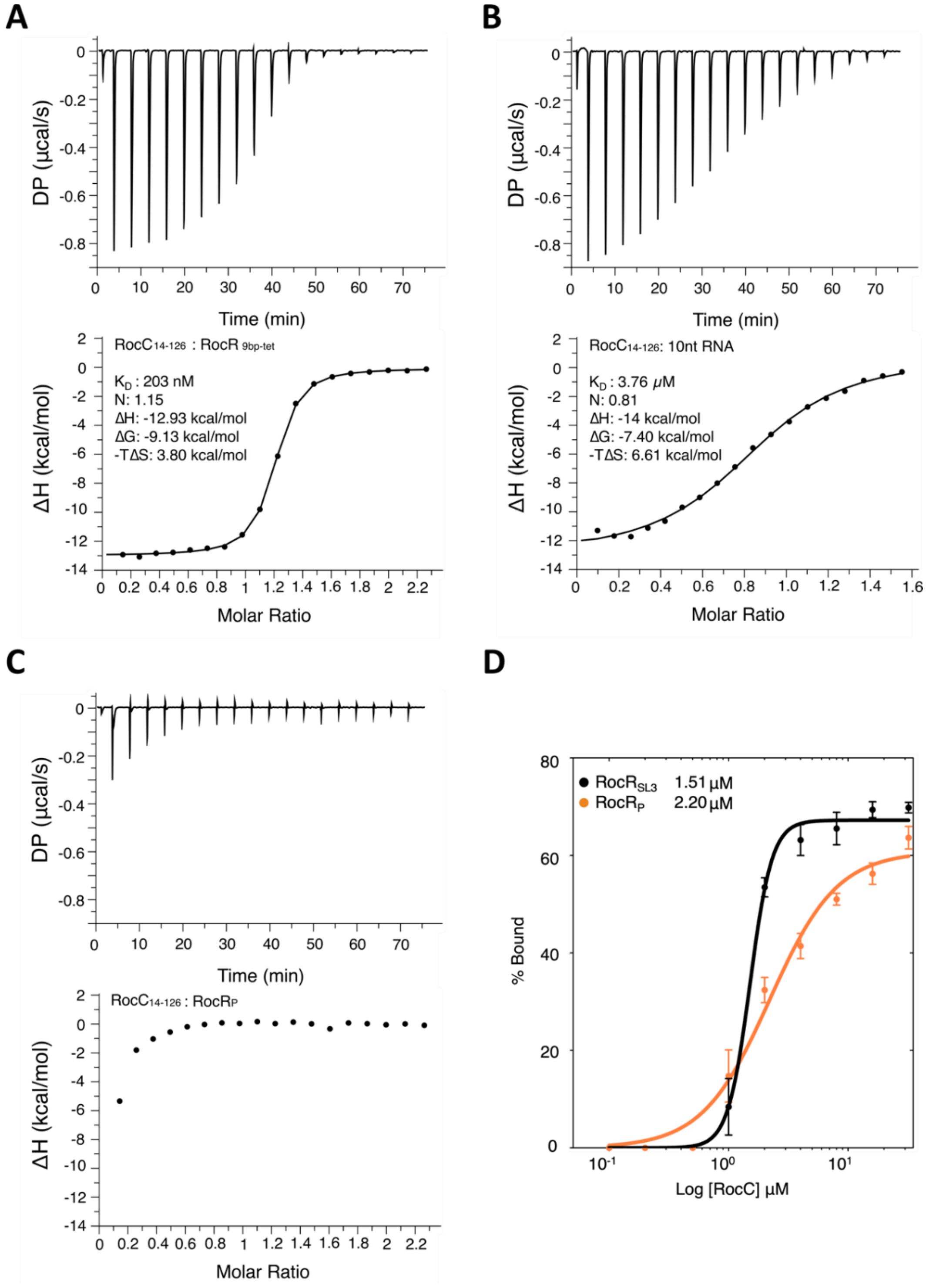


Figure 5-4. Measurement of binding interactions between RocC$_{14-126}$ and either RocR$_{9bp-tet}$, a single stranded RNA, or RocR$_{P}$.

(A) ITC analysis of RocR$_{14-126}$ with RocR$_{9bp-tet}$. (B) ITC analysis of RocR$_{14-126}$ with a 10 nucleotide single-stranded RNA. (C) ITC analysis of RocR$_{14-126}$ with RocR$_{P}$. (D) EMSA analysis of RocR$_{14-126}$ with T4PNK treated RocR$_{P}$.

possible if the RNA strand is in a A-form geometry (Figure 4-8B). To test the hypothesis that the RocC helical N-cap motif could impose a specificity for hairpin-containing RNA partners compared to single-stranded RNA, we used ITC to compare the interactions of RocC$_{14\text{-}126}$ to either RocR$_{9bp\text{-}tet}$ or a 10 nucleotide single-stranded RNA corresponding to just the RNA region directly in contact with RocC in the crystal structure (Figure 5-3A, B). The results confirmed a tight 1:1 interaction between RocC and RocR$_{9bp\text{-}tet}$ and furthermore revealed that the interaction is largely enthalpy-driven. RocC also bound the ssRNA with a 1:1 stoichiometry, albeit with an affinity that was ~19-fold weaker than for the stem-loop structure. In this case, the binding was still enthalpically-driven, however, the entropic cost of binding was much higher, consistent with the idea that the single-stranded RNA is able to make the same interactions with RocC, albeit with a higher entropic penalty due to the structural restraints imposed by the binding interaction on the flexible ssRNA (Figure 5-5 A, B).

The tight packing of the 3' nucleotide of RocR against RocC as well as the interaction of the 3'-hydroxyl with the penultimate phosphate suggests specific recognition of a terminal nucleotide with a 3'-hydroxyl group. To test this idea, we assessed the interactions of RocR$_{SL3}$ containing a terminal 3'-phosphate with RocC by ITC, demonstrating the critical importance of the chemical structure of the terminal nucleotide for RocC recognition (Figure 5-3C). The 3' phosphate group in RocR$_{P}$ was removed by treatment with T4PNK, which contains a 3' phosphatase domain in addition to its kinase domain. Binding of this RNA was tested using EMSA. The EMSA demonstrated the recovery of binding affinity of RocR$_{P}$, which corroborated the importance of the 3' hydroxyl group for target recognition of RocC (Figure 5-3D).

5.3.3. ProQ/FinO domain proteins exhibit specificity for the length of the terminator 3' tail

ProQ/FinO family members exhibit profound differences in their ability to recognize different RNAs *in vivo*. Certain proteins, such as RocC and FinO, only have one biological partner whereas other family members, such as ProQ and NMB1681, can bind a range of RNAs. The structures of the ProQ/FinO domains of these proteins, in particular, their 3' nucleotide binding pockets and N-cap motifs, are well conserved between these proteins. Previous work demonstrated that FinO, ProQ and NMB1681, like RocC, all can bind transcription terminator structures (Attaiech et al., 2016; Chaulk et al., 2011; Jerome & Frost, 1999). Indeed, modelling of FinO and

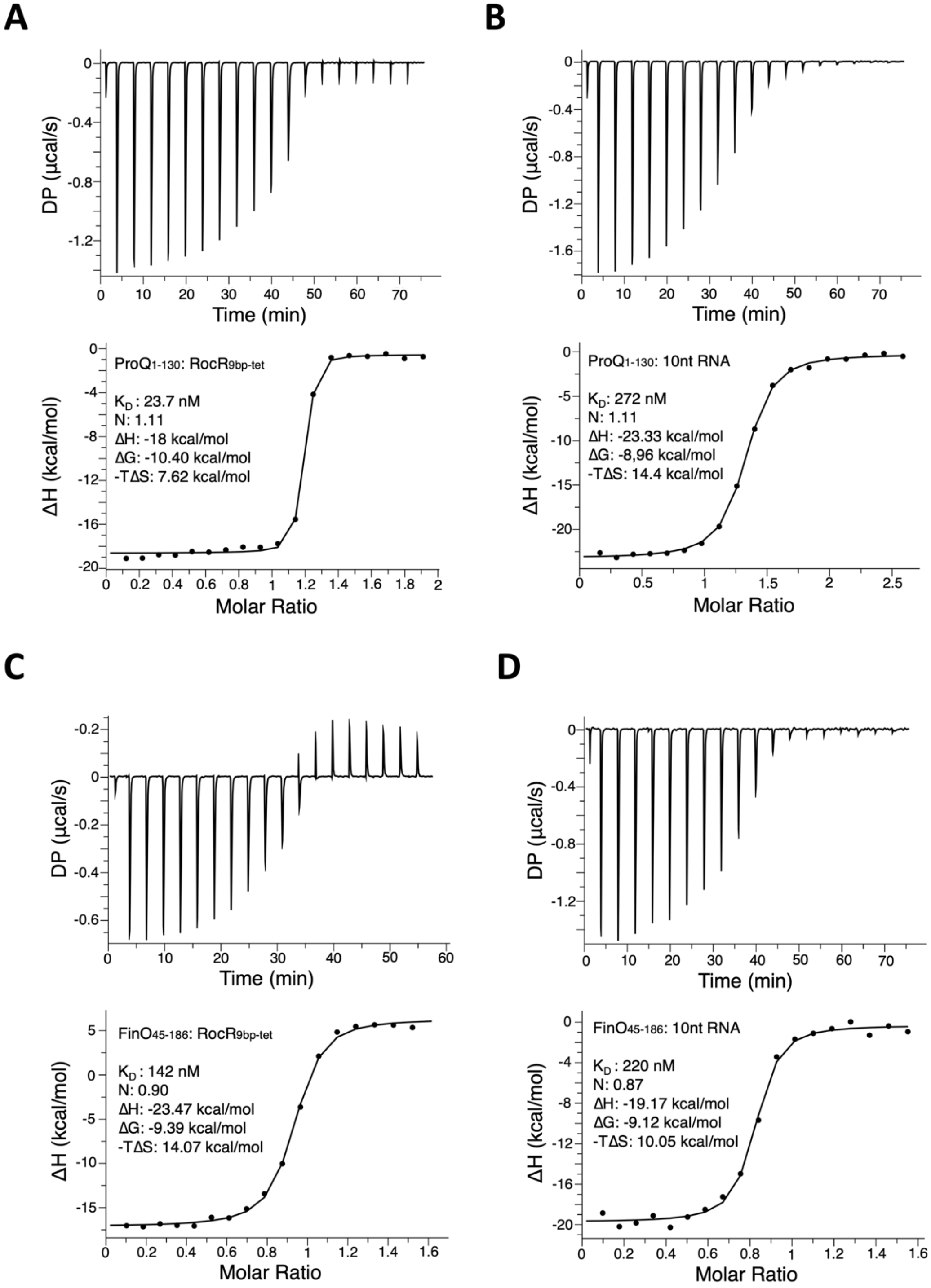


Figure 5-5. Measurement of binding interactions between ProQ/FinO proteins and either $\text{RocR}_{\text{9bp-tet}}$ or a single stranded RNA.

(A) ITC analysis of $\text{ProQ}_{1\text{-}130}$ with $\text{RocR}_{\text{9bp-tet}}$. (B) ITC analysis of $\text{ProQ}_{1\text{-}130}$ with a 10 nucleotide single-stranded RNA. (C) ITC analysis of $\text{FinO}_{45\text{-}186}$ with $\text{RocR}_{\text{9bp-tet}}$. (D) ITC analysis of $\text{FinO}_{45\text{-}186}$ with a 10 nucleotide single-stranded RNA.

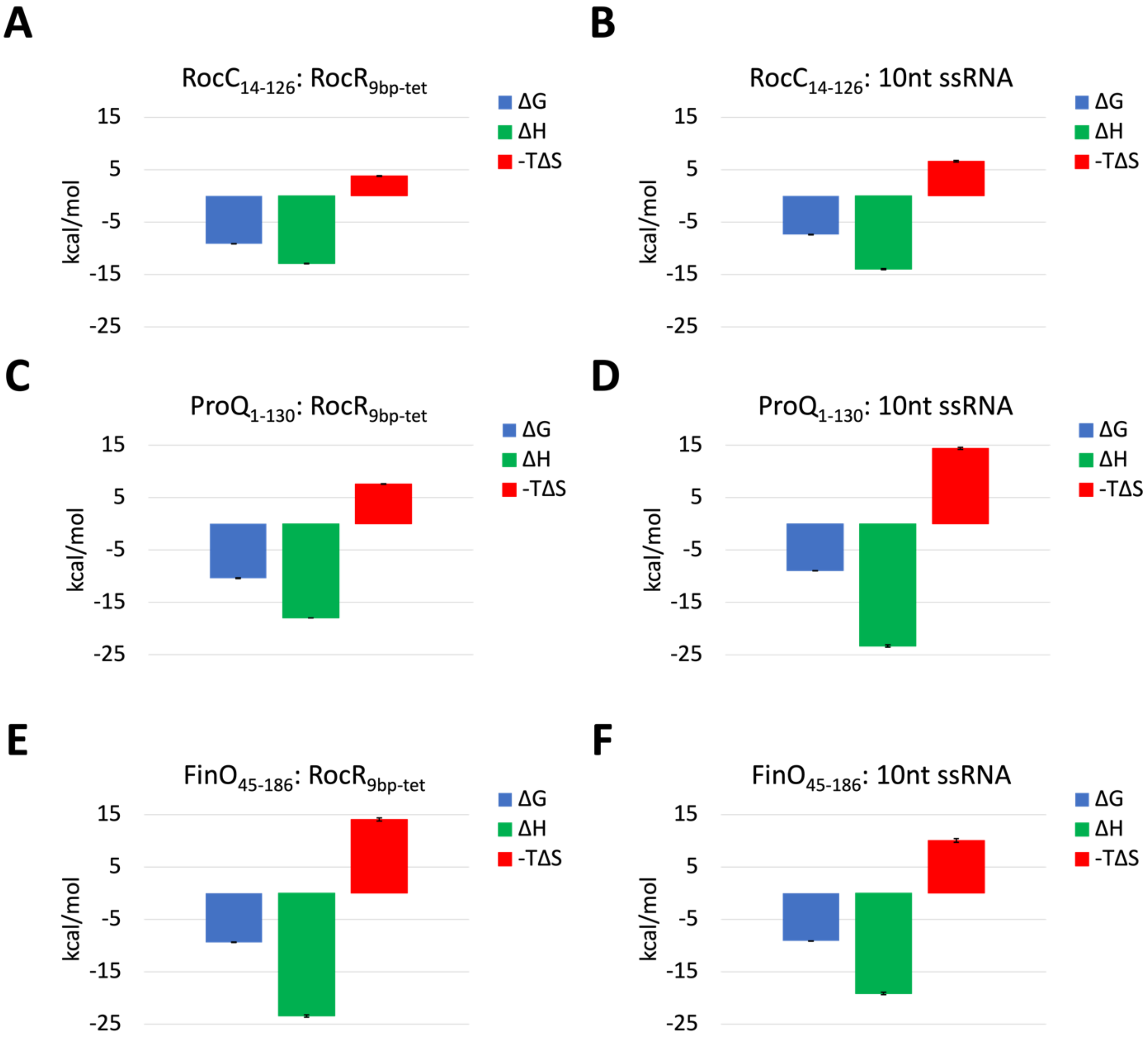


Figure 5-6. Thermodynamic data between the isolated ProQ/FinO domains and either RocR$_{9bp\text{-}tet}$ or a single stranded RNA.

(A) Thermodynamic data of RocC$_{14\text{-}126}$ with RocR$_{9bp\text{-}tet}$. (B) Thermodynamic data of RocC$_{14\text{-}126}$ with a 10 nucleotide single-stranded RNA. (C) Thermodynamic data of ProQ$_{1\text{-}130}$ with RocR$_{9bp\text{-}tet}$. (D) Thermodynamic data of ProQ$_{1\text{-}130}$ with a 10 nucleotide single-stranded RNA. (E) Thermodynamic data of FinO$_{45\text{-}186}$ with RocR$_{9bp\text{-}tet}$. (F) Thermodynamic data of FinO$_{45\text{-}186}$ with a 10 nucleotide single-stranded RNA.

ProQ suggests that the ProQ/FinO domains of each of these proteins could bind to hairpin-3' tail RNAs very similarly to how RocC binds the RocR terminator. We hypothesized that part of the difference in specificity of these different proteins might be their ability to bind RNAs with different lengths of 3’ single-stranded tails. To test this, we created a set of model terminators based on RocR_{SL3} with 3, 5, or 8 nucleotide tails and measured the affinity of the ProQ/FinO domains of RocC, ProQ and FinO for these RNAs by FP (Figure 5-6). All the ProQ/FinO domains bound at least one of the test RNAs with high affinity ($K_D < 1$ μM). In agreement with the EMSA results, RocC showed a dramatic specificity for the 5-nucleotide tail, and only interacted weakly with the 3- or 8-nucleotide tails ($K_D > 20$ μM). Similarly, the ProQ/FinO domain of ProQ showed a strong binding to the 5-nucleotide tail RNA, and less binding to the other RNAs. In contrast, the ProQ/FinO domain of FinO showed less difference in binding specificity between the different RNAs, with only a weak preference for the 5- and 8-nucleotide tail lengths compared to the 3-nucleotide tail.

5.4. Discussion

5.4.1 RocC:RocR provides a model for the recognition of intrinsic transcriptional terminators by proteins with a ProQ/FinO domain

The structure of the RocC:RocR_{SL3} complex presented in the chapter 4 reveals a mechanism where RocC specifically binds its natural partner RocR by interacting with its terminator, i.e. the hairpin of SL3 and its 3’ single-stranded polypyrimidine tail. The relevance of this structure to RocC:RocR binding in solution and *in vivo* is supported by extensive mutagenesis data. This structure also provides a model to understand RNA recognition by other ProQ/FinO domain proteins. The pocket that binds the 3’ nucleotide of the terminator is particularly well conserved in the ProQ/FinO family. The position of the N-cap motif is highly conserved in the ProQ/FinO family, and Arg75 is almost completely conserved (80% identity) with the major substitution being lysine. The structure explains previous biochemical and mutagenesis data on FinO as well as ProQ. Site-specific protein-RNA crosslinking identified residues in and around the conserved 3’ nucleotide binding pocket (Lys125, Lys165), as well as the N-cap (Arg121) as key contact points for FinO (Ghetu et al., 2002). Likewise, bacterial 3-hybrid experiments identified the same surfaces on the ProQ/FinO domain of ProQ as important to recognize transcriptional terminator structures (Pandey et al., 2020). Ribonuclease footprinting revealed dramatic protection of the FinP 3’ tail by FinO, as well as protection of the first 3-4 nucleotides within the stem, precisely the same region that is bound by RocC in the RocC:RocR structure (Arthur et al., 2011). Interestingly, it was previously

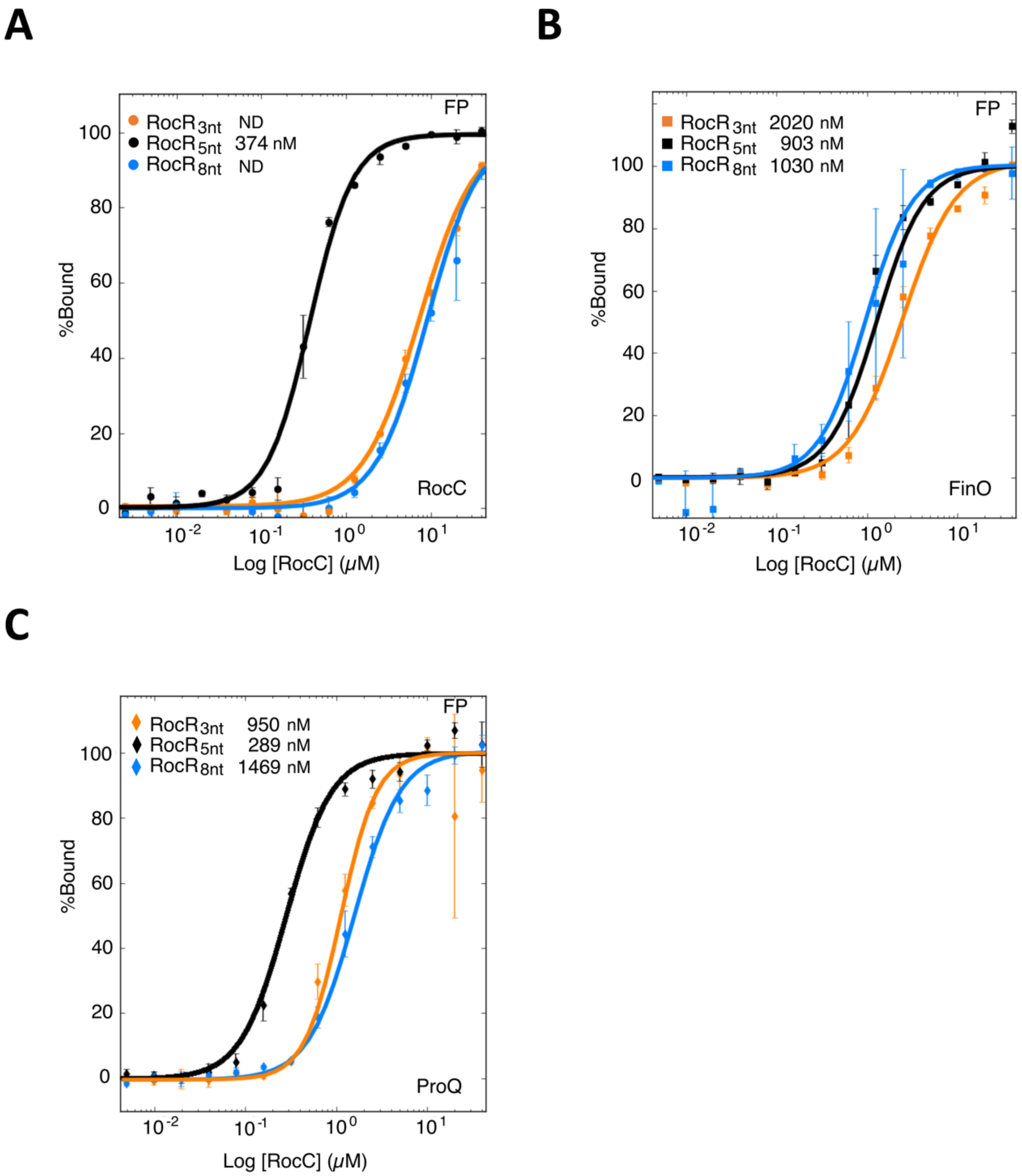


Figure 5-7. Specificities for 3’ tail length among FinO domain-containing proteins.

(A, B, C) FP binding assay for various tail lengths of 5’ FAM-labeled RocR$_{SL3}$ with RocC$_{14-126}$ (B), with FinO$_{45-186}$ (C), or with ProQ$_{1-130}$. ND indicates mutants where RNA binding was too weak to determine a K$_D$.

shown that the recognition of FinP by FinO is strongly dependent on the chemical structure of the

3’ nucleotide (Arthur et al., 2011), and we demonstrate here that phosphorylation of the 3’-nucleotide of RocR also abolishes RocC binding (Figure 5-3C).

The ProQ/FinO family members that have been studied in detail (FinO, RocC, ProQ and NMB1681) all bind transcriptional terminator structures but with varying degrees of specificity. While FinO and RocC bind to just one or two physiological partners, ProQ and NMB1681 bind to dozens of different sRNAs. Our structure of the RocC:RocR$_{SL3}$ complex suggests a mechanism for the specific recognition of the hairpin as well as for the 3’ polypyrimidine tail. The duplex portion of the hairpin is recognized on one strand by the N-cap motif. The N-cap motif plays a key role in interaction. The structure showed the Ser70 and Ser 72 on the end of the N-cap clamps phosphate oxygens (Figure 4-8A, B). Sequence alignment of the ProQ/FinO domain also showed that Ser70 and Ser72 residues are highly conserved with serine and threonine, emphasizing the importance of the hydroxyl group in this position. Interestingly, the mutation of the individual Ser70 and Ser72 showed minor influences *in vitro* binding and *in vivo* function. Double mutations, on the other hand, lost binding to RNA and the weaker binding correlated to reduce transformability repression. Serine and threonine residues often contribute to nucleic acid interactions through phosphate backbone contacts (Luscombe et al., 2001). Specificity for the duplex is likely conferred through recognition of the A-form geometry of the contiguous phosphate groups on a single strand, rather than through a direct recognition of both strands of the hairpin. This is supported by our finding that while an ssRNA can bind RocC, its binding is entropically disfavored compared to the hairpin form (Figure 5-4). The lack of any direct interactions between the base-pairs of the stem and RocC is consistent with the lack of apparent sequence specificity of recognition of this domain. We found that replacement of the G-C base-pairs at the base of the stem only resulted in a small (~2-fold) reduction in binding affinity and previous results with FinO also indicate that this protein does not recognize specific sequences at this position (Jerome & Frost, 1999). The 3’ polypyrimidine strand adopts a hook-like structure, stabilized by hydrogen bonding between the penultimate phosphate and the terminal 3’-hydroxyl (Figure 4-8C). In this conformation, the strand tracks into the 3’ nucleotide binding pocket. While it is difficult to definitively model side chain – base hydrogen bonding interactions at this resolution, the structure does suggest some possibilities that may indicate specificity for pyrimidines within the tail. Highly conserved Arg75 participates in RNA backbone recognition, but may also hydrogen bond with the 5’-most uridines of the polypyrimidine tail. The next 2 nucleotides are packed against α7 in a manner that may sterically favour pyrimidines over purines, and residues such as Lys119 and Asn115 may provide sequence specificity through hydrogen bonding interactions with these bases. Consistent with this idea, mutations at these residues reduce binding affinity and, partially, transformation repression efficiency. The uridine base of the terminal 3' nucleotide packs against Thr82 and mutation of this residue does significantly reduce transformation repression efficiency.

Despite the highly conserved 3' pocket and N-cap motif in FinO and ProQ, the isolated ProQ/FinO domain of FinO and ProQ showed different behaviour toward ssRNA in ITC. FinO does not strictly discriminate between stem-loop structures and ssRNAs. ProQ showed a 10-fold weaker interaction with ssRNA. At this point, we cannot explain why FinO apparently does not discriminate between the hairpin-containing terminator structure and the ssRNA, which is in contrast to our previous findings using FinP SLII and related ssRNAs. However, we do note that in the FinO binding experiment we did note a small positive heat of mixing in the saturated part of the titration. If this corresponds to a heat of mixing due to an imbalance in solution components, then it is possible that the measured values may underestimate the molar heat of binding and the affinity for the hairpin-containing RNA.

In all cases, the binding reaction was driven mainly by enthalpy and was entropically unfavorable (Figure 5-5). Enthalpically favourable reactions are commonly observed for protein/nucleic acid interactions that are stabilized mainly through hydrogen bond and van der Waals interactions (Prozeller et al., 2019) (Figure 4-8A, B, C). Our thermodynamic analysis of RocC$_{14\text{-}126}$:RNA interactions showed that the most of the reduction in binding affinity and ΔG was due an increase of +2.81 kcal/mol in $-T\Delta S$, indicating a large reduction in entropy when comparing binding of the hairpin-containing RNA to the ssRNA. Entropy changes are linked to hydrophobic interaction and conformational constraints (Prozeller et al., 2019). Considering that protein/RNA interactions are mainly driven by hydrogen bond and van der Waals forces, the negative entropy changes are likely strongly influenced by the conformational constraints imposed on the ssRNA upon binding to RocC$_{14\text{-}126}$. This entropic penalty is less when considering the hairpin RNA which is already held in the A-form conformation that is favourable for binding.

Our results also suggest that different ProQ/FinO domains have different specificities for the length of the 3' ssRNA tail (Figure 5-6). Biologically, both FinO and RocC bind RNAs with 5-nucleotide ssRNA tails, while ProQ tends to bind sRNA targets with shorter 4-nucleotide tails

(Holmqvist et al., 2018; Stein et al., 2020). Biochemically, we find that the isolated ProQ/FinO domains of RocC, FinO and ProQ all have differing tail length specificities. RocC binds very specifically to its 5-nucleotide tailed RNA target. ProQ also preferentially bound to the 5 nucleotide-tailed RocR compared to versions with either 3- or 8-nucleotide tails, while FinO showed little preference for the different tail lengths. Our structure would predict a minimal ssRNA tail length of 3 nucleotides, corresponding to the portion of the tail which must contour into the nucleotide binding pocket. In our structure, the 5' nucleotide forms a U-U mispair with the first nucleotide of the 3' poly-pyrimidine tract, where the 5' nucleotide is packed against α5, however it is unclear if this mispair would be possible in the case of a longer RNA extended at the 5' end. It is possible that different ProQ/FinO domain proteins all bind the 3' strand in a way that is similar to what is observed in the RocC/RocR complex, however differences in how much the 5' region can pair with the 3' region might be determined by interactions with residues from α 5 and possibly α7.

Chapter 6.

General discussion

6.1. Overall summary

Phylogenetic studies have revealed that ProQ/FinO family proteins are widely distributed in proteobacteria and RNA-seq demonstrated that they regulate genes associated with structured sRNAs and mRNAs (Attaiech et al., 2016). ProQ/FinO domains play a crucial role in RNA binding; however, the molecular details on binding remained unclear. RocC is one of the ProQ/FinO family in *L. pneumophila*, essential for *L. pneumophila* to repress natural transformation under normal growth conditions. RocC regulates the development of competence in association with its sRNA, RocR, which consists of three stem-loops: SL1, SL2, and SL3. A previous study showed that the ProQ/FinO domain of RocC (RocC$_{1\text{-}126}$) binds SL3, and SL1 and SL2 interact with target mRNAs related to competence development through the 5' UTR of the mRNAs (Attaiech et al., 2016). In this study, we tried to understand the minimal determinants of SL3 and the ProQ/FinO domain to bind each other using EMSAs. This knowledge contributed to obtaining diffracting crystals and determining a crystal structure of RocC$_{14\text{-}126}$ (ProQ/FinO domain) in complex with RocR$_{9bp\text{-}tet}$. The crystal structure analysis revealed for the first time how a ProQ/FinO domain recognizes an intrinsic transcriptional terminator in a highly specific manner. Structure-guided mutagenesis *in vitro* and *in vivo* demonstrated the importance of specific structural features for RNA binding and binding and the repression of competence.

Initial binding assays with RocC and RocR truncation mutants provided insights into minimum constructs for structural studies. As suggested by a previous EMSA study, we focused on the binding of the ProQ/FinO domain and SL3 (RocR$_{SL3}$). The flexible N-terminal extension in RocC$_{1\text{-}126}$ was truncated (RocC$_{14\text{-}126}$ and RocC$_{24\text{-}126}$) and tested the impact in binding. The binding affinity of RocC$_{24\text{-}126}$ with RocR$_{SL3}$ was too weak to measure in these initial assays, and was dramatically weaker than the binding of either RocC$_{1\text{-}126}$ or RocC$_{14\text{-}126}$ for SL3 (Figure 3-5A). These EMSA results concluded that RocC$_{14\text{-}23}$ is essential for tight binding, and that RocC$_{14\text{-}126}$ would be a promising candidate for future structural studies. Investigation of RocR$_{SL3}$ truncation mutations using EMSA and FP confirmed previous data that the intrinsic terminator should have a minimal 3' tail length and stem base. A remarkable discovery was that RocC particularly requires a 5 nucleotide 3' tail, unlike ProQ and FinO. We crystallized RocC$_{14\text{-}126}$:RocR$_{9bp\text{-}tet}$ and determined the structure at 3.2 Å to reveal the secret of the 5 nucleotide tail specificity.

Determination of the crystal structure elucidated why RocC requires a 5 nucleotide tail and small duplex RNA stem. The 5 consecutive phosphate groups on the 3' side of the stem base ($G_{19}G_{20}C_{21}C_{22}C_{23}$) is contacted directly by a conserved α-helical N-cap motif of RocC (Figure 4-8A, B). Within the 3' single-stranded tail, only the terminal two nucleotides, $C_{27}U_{28}$, directly contacted RocC through interactions with the highly conserved 3' nucleotide binding pocket. The next two residues immediately 5' to these nucleotides, $U_{25}U_{26}$, serves as a linker to place $C_{27}U_{28}$ in the 3' binding pocket (Figure 4-8A, C). The 2' and 3' hydroxyl groups of U28 forms hydrogen bonds with the main chain amine groups in a type II β-turn-α-helix motif that is highly conserved in the ProQ/FinO family.

Based on structural analysis, we hypothesized that the 3' hydroxyl group of last ribose sugar of the single-stranded tail, and the A-form helical structure of the hairpin, are the primary determinants recognized by RocC. ITC demonstrated that 3' phosphorylation of the 3' terminal residue of $RocR_{SL3}$ blocked $RocC_{14\text{-}126}$ binding, suggesting 3' hydroxyl group recognition is key for specific binding (Figure 5-3). A-form helical structure recognition was tested with a single-stranded 10 nucleotide RNA using ITC. The binding affinity of $RocC_{14\text{-}126}$ for this ssRNA was reduced 19-fold compared to $RocR_{SL3}$. Even though both interactions were enthalpically-driven, the ssRNA showed a higher entropic penalty compared to $RocR_{SL3}$. The result from ITC suggested that the lower stem was required to conformationally restrict the A-form helix of 3' side despite no direct contact on 5' side.

In conclusion, our comprehensive studies on $RocC_{14\text{-}126}$ in complex with a high-affinity intrinsic transcriptional terminator revealed the mechanistic principles underlying this recognition. The highly conserved nature of the binding surfaces and RNA binding data from other ProQ/FinO proteins suggest key features of this mechanism are likely largely conserved across the family.

6.2. Initial target filtering in the ProQ/FinO domain

It is well known that ProQ/FinO family is widely distributed across proteobacteria (Attaiech et al., 2016). They often contain either N-terminal or C-terminal extensions which are usually often poorly conserved and predicted to be highly disordered. Binding studies with truncation mutants defined ProQ/FinO domains as an RNA binding anchor that recognize intrinsic terminators (Arthur et al., 2011; Chaulk et al., 2010; Jerome & Frost, 1999). Cross-linking or immunoprecipitation followed by RNA-sequencing enabled the detection of target RNAs for several FinO/ProQ family members. Application of these technologies showed that certain ProQ/FinO family members, such as RocC and FinO (Attaiech et al., 2016; El Mouali et al., 2021), could be highly selective, binding only one or two RNA targets, while others, such as ProQ and NMB1681, could bind dozens of different RNA targets (Bauriedl et al., 2020; Smirnov et al., 2016). Even though our structure of the RocC:RocR complex revealed the secret of the high target selectivity of RocC, it remained unclear how ProQ, which contains a high degree of structural similarity, binds a diverse array of RNA targets. Our structural analysis cannot clarify what structural feature causes this difference; however, additional binding assays using FP and ITC in this thesis can provide more clues to understanding this conundrum.

We demonstrated that RocC$_{14\text{-}126}$ in the absence of its C-terminus is sufficient to achieve high selectivity. Comprehensive data presented in this thesis (Biochemically in Chapter 3, Structurally in Chapter 4, Biophysically in Chapter 5) suggested that high selectivity is rooted in ProQ/FinO domain of RocC and Figure 6-1 showed models to explain why ProQ/FinO domain of RocC requires certain tail lengths of RNA.

Unlike RocC and FinO, ProQ binds a large RNA target suite and this recognition is also rooted in the ProQ/FinO domain. Unlike the ProQ/FinO domain of RocC, which shows a high degree of discrimination toward tail length and A-form RNA, ProQ$_{1\text{-}130}$ showed generous binding to a range of RNA constructs. FP and ITC analyses with different RNA constructs using RocC$_{14\text{-}126}$ and ProQ$_{1\text{-}130}$ in chapter 5 demonstrated the promiscuous nature of ProQ$_{1\text{-}130}$. FP studies with three different tail lengths of RocR$_{SL3}$ showed tight binding to all three tails although the tightest was RocR$_{5nt}$ (Figure 5-6C), supporting previous EMSA data that the minimal ProQ target consists of a 2 base-pair stem with a 4 nucleotide 3' tail (Stein et al., 2020). The importance of 3' tail length was tested using the *malM* mRNA 3' UTR (a validated *in vivo* ProQ target) where the polyuridine tail length was varied from one to nine uridines. Tails of one to three uridines failed to bind; however, the ProQ/FinO domain of ProQ showed tight binding to RNA with four uridines (K_D: 27 nM) to nine uridines (K_D: 5.1 nM). ProQ/FinO domain of ProQ also binds to ssRNA. Even though

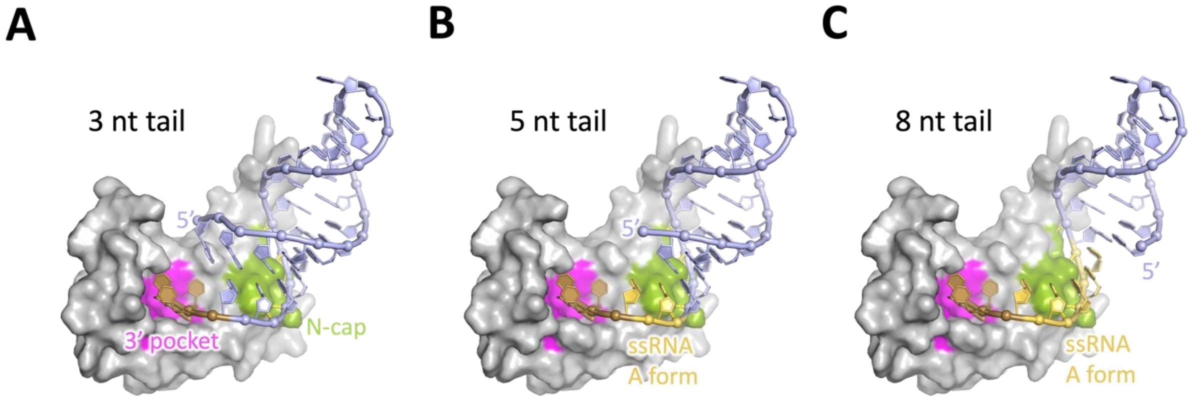


Figure 6-1. Models for a different mode of binding depending on 3' tail length.

(A-C) Models for a 3, 5, and 8 nucleotide 3' tail with ProQ/FinO domain of RocC. Magenta and light green on protein surfaces indicated a 3' binding pocket and N-cap motif, respectively. Yellow and light purple on RNA showed single-stranded and double-stranded, respectively.

the ITC binding affinity of $ProQ_{1-130}$ with ssRNA was 10-fold weaker compared to $RocR_{SL3}$, it is interesting to note that ProQ binds much tighter to these RocR-derived RNAs than RocC.

In addition to the ProQ/FinO domain, ProQ also contains a C-terminal Tudor domain connected to the ProQ/FinO domain through a flexible linker and there is some evidence that all three regions of ProQ contribute to RNA interactions. For example, *cspE* mRNA 3' UTR is a known ProQ target with a 4 polyuridine 3' tail. Both the isolated ProQ/FinO domain and full-length ProQ bind *cspE* 3' UTR however, only full-length ProQ can bind *cspE* 3' UTR mutants with long extensions to the 3' tail, suggesting that the ability of ProQ to bind multiple RNA targets might involve the additional C-terminal domain. Direct evidence for interactions of the C-terminal domain with RNA come from HDX experiments, which showed all parts of ProQ – the ProQ/FinO domain, linker, and Tudor domains - all change their solvent accessibility when bound to RNA (Gonzalez et al., 2017). In addition, a comparison of ProQ protected surfaces when bound to two different RNA targets, SraB and *cspE* mRNA 3' UTR, revealed that different regions were protected in the two different complexes. In conclusion, the large target suite of *E. coli* ProQ may result from the promiscuous binding of its ProQ/FinO domain, together with the assistance of the linker and Tudor domain.

6.2. Additional selectivity could involve regions outside the ProQ/FinO domain

Our FP and ITC assays of the ProQ/FinO domain from FinO ($\text{FinO}_{45\text{-}186}$) showed interesting results that suggest that FinO binding specificity might not only be conferred by ProQ/FinO domain, but may also involve the flexible N-terminus. FinO tail length specificity was previously investigated with SLII of FinP using EMSA (Jerome & Frost, 1999). This work demonstrated a strong preference of binding to an SLII with a 6 nucleotide tail compared to SLII variants with either a 4 or 2 nucleotide tail. Our FP and ITC affinity measurements using just the core ProQ/FinO domain ($\text{FinO}_{45\text{-}186}$) showed little affinity differences for RocR SL3 variants with either 3, 5 or 8 nucleotide tails (Figure 5-6B). Moreover, these experiments also show that, unlike RocC, there is little difference in binding affinity between SL3 and the 10 nucleotide ssRNA. This is surprising, given the fact that FinO is highly specific for FinP *in vivo*, while ProQ is apparently much more promiscuous in terms of its range of *in vivo* RNA partners (Figure 5-4, Table 5-1B). RNA sequencing has recently confirmed the high specificity of FinO *in vivo*, revealing only two target RNAs in *E. coli*, FinP and RepX, which have similar structures (El Mouali et al., 2021).

One difference between the previous EMSA and our current FP assay is that EMSA used a natural binding target, SLII, and FP used RocR_{SL3}. While we cannot rule out a role for these sequence differences, it is important to note that previous analysis of the effects of SLII mutations on FinO binding affinity did not uncover significant RNA sequence specificity(Jerome & Frost, 1999). The major difference might be the FinO construct used in EMSA and FP assays. The EMSA used GST tagged full-length FinO (GST-FinO) and FP and ITC used $\text{FinO}_{45\text{-}186}$, which lacks the flexible region N-terminal to the core ProQ/FinO domain. Could it be that this flexible N-terminus provides additional binding specificity? It is interesting to note that in a comparison of the affinities of N-terminal truncation mutants of FinO, it was found that $\text{FinO}_{45\text{-}186}$ actually binds SLII 25-fold tighter than full-length FinO, suggesting a role for the N-terminus. An example of higher selectivity from an unstructured tail is found in Hfq protein. Hfq forms a homohexameric ring to perform its biological role as an RNA chaperone. The monomer has a conserved Sm-like fold with unstructured N- and C-terminal flanking regions. The N-terminus is quite short (10-20 nucleotides), however, the C-terminus length can vary between species (*E. coli*: 38 amino acids, *B. subtilis*: 9 amino acids). This unstructured part is vital for chaperone activity and RNA selectivity (Santiago-Frangos et al., 2017). Sequence alignment revealed a conserved acidic part of the C-terminus. The interaction of the acidic tail and the arginine rich RNA binding surface on the rim of Hfq was demonstrated using fluorescence anisotropy. The acidic C-termius competed with the target RNA to bind through the arginine-rich region. This result suggested that the acidic region of the C-terminus mimics RNA in an autoinhibitory mechanism to increase the specificity of Hfq (Santiago-Frangos et al., 2017). The interaction of FinO$_{1\text{-}61}$ (N-terminus) with FinP was determined using EMSA to be 710 nM, 3.6-fold weaker than full-length FinO (Ghetu et al., 1999). Three residues within the FinO N-terminal region, Lys37, Lys40, and Lys 42, were crosslinked with SLII of FinP and gelFRET revealed the close proximity of Lys37, Lys42 with the base of the SLII stem (Ghetu et al., 2002). Taken together, these results support the idea that the structurally flexible N-terminus of FinO may directly contact RNA to increase the target selectivity and this idea should be tested with further experiments.

6.3. Possibility of a different binding mode

Our structural study revealed that two distinct surfaces are employed by the ProQ/FinO domain to bind RNA; the N-cap and the 3’ tail binding pocket (Figure 4-8A, B, C). Mutational studies demonstrated that point mutations in both the N-cap and the 3’ tail binding pocket cause defects in binding and biological function (Chapter 5).

NMB1681 in *N. meningitidis* lacks N- and C-terminal extensions and as such is essentially just a minimal ProQ/FinO domain. This small protein can bind hundreds of sRNA and mRNAs to respond to cellular stress (Bauriedl et al., 2020). Like other ProQ/FinO domains, this protein contains the positively charged 3’ tail binding pocket, however, it also has a more positive charge distribution on the convex side (Figure 6-2A, B) (Olejniczak & Storz, 2017). Sequence alignment of the N-cap regions among the five determined structures shows that NMB1681 has an unusual N-cap that is not conserved in the other structures (Figure 6-2C). The highly conserved arginine corresponding to Arg75 in RocC is alanine in NMB1681 (Figure 6-2D, E). NMB1681 however does have an arginine one residue N-terminal to this position which may play a role in RNA binding (Figure 6-2E). Perhaps even more striking is the exchange of the conserved N-capping serine residue for an aspartic acid. This residue does cap the helix in NMB1681, however the negative charge of this residue suggests that interactions with RNA targets will be different for this region than observed in RocC (Figure 6-2D).

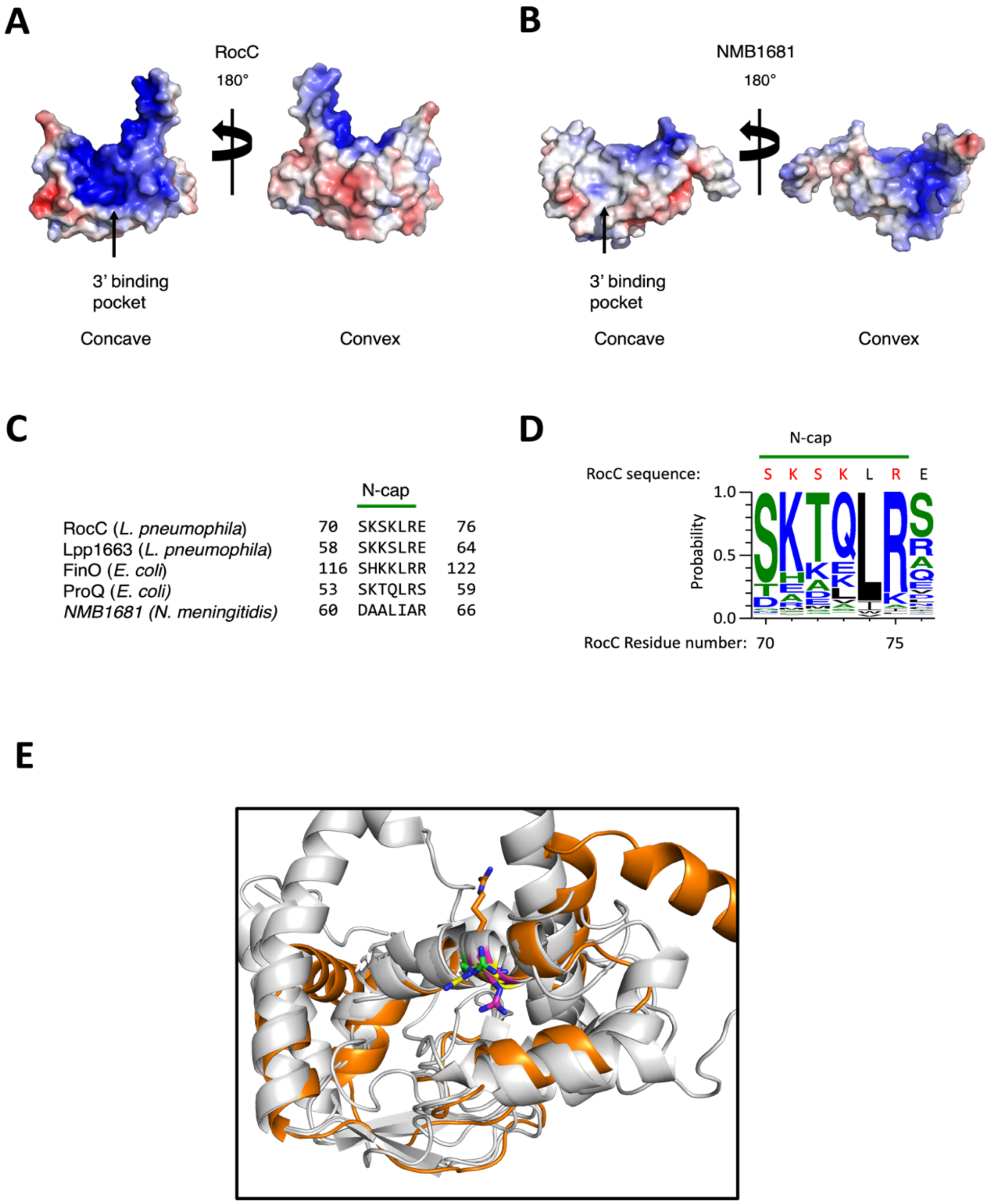


Figure 6-2. ProQ/FinO domain of NMB1681 showed a difference compared to other reported structures.

Electrostatic potential surfaces of (A) ProQ/FinO domain of RocC (B) NMB1681. (C) Sequence alignment of N-cap motif with known structures. (D) WebLogo of N-cap region (Figure 4-12B). (E) NMB1681 (orange) structural alignment with RocC, ProQ, and FinO. Orientations of Arg66 (orange, NMB1681), Arg75 (magenta, RocC), Arg58 (green, ProQ), Arg121 (yellow, FinO) are shown.

6.4. Implications of the RocC:RocR structure for RNA remodeling

One of the most intriguing properties of ProQ/FinO domain RNA chaperones is their ability to directly facilitate RNA-RNA association. This has been best studied for FinO, where the protein has been demonstrated to facilitate sense-antisense pairing between FinP and *traJ* mRNA *in vitro* (Arthur et al., 2003; Biesen & Frost, 1994; Ghetu et al., 2000; Gubbins et al., 2003; Sandercock & Frost, 1998) (Figure 6-3). FinO can also enhance interactions between the minimal transcriptional terminator SLII of FinP and its complementary region of *traJ* and it can also catalyze strand exchange between duplex and ssRNAs, implying that FinO can destabilize the double-stranded nature of the bound SLII hairpin. Sense-antisense recognition is thought to proceed via kissing interactions between complementary loop regions of the RNAs with subsequent unwinding of internal hairpins and inter-strand duplexing (Mark Glover et al., 2015). However, how FinO could at once bind to a hairpin RNA while destabilizing its base pairing was unclear. The RocC:RocR structure reveals that RocC only binds the 3' strand of the hairpin, leaving the 5' side unencumbered so that it could be peeled away to allow duplexing with a complementary RNA.

The exact role of other ProQ/FinO domain proteins in facilitating RNA-RNA association is less clear. Many of these proteins, such as RocC and ProQ, mediate sRNA – target RNA recognition where there is only limited base pairing between the two RNAs (Figure 6-3). Most ProQ/FinO domain RNA partners identified to date are highly structured, and it is intriguing to speculate that the intrinsically disordered regions outside of the ProQ/FinO domain may be critical for chaperone activity. This hypothesis is supported by the observation that both the FinO_N domain of FinO (N-terminal to the ProQ/FinO domain) and the C-terminal domain of RocC (C-terminal to the ProQ/FinO domain) were shown to be essential to their function. Deletion of the flexible FinO_N domain abrogated duplexing and strand exchange activities but did not reduce RNA binding activity of FinO. This mutant was also unable to repress conjugation which suggests that this domain is critical for the RNA chaperone activities of FinO *in vivo* (Arthur et al., 2003). In a similar manner, deletion of the C-terminal domain of RocC does not impair its binding to RocR or RocR stability but causes a loss of the post-transcriptional repression of the mRNA targets (Attaiech et al., 2016) (Figure 6-3). More studies are needed to better understand the potential dynamic interactions between the ProQ/FinO domain, the associated domains, and their RNA partners and targets.

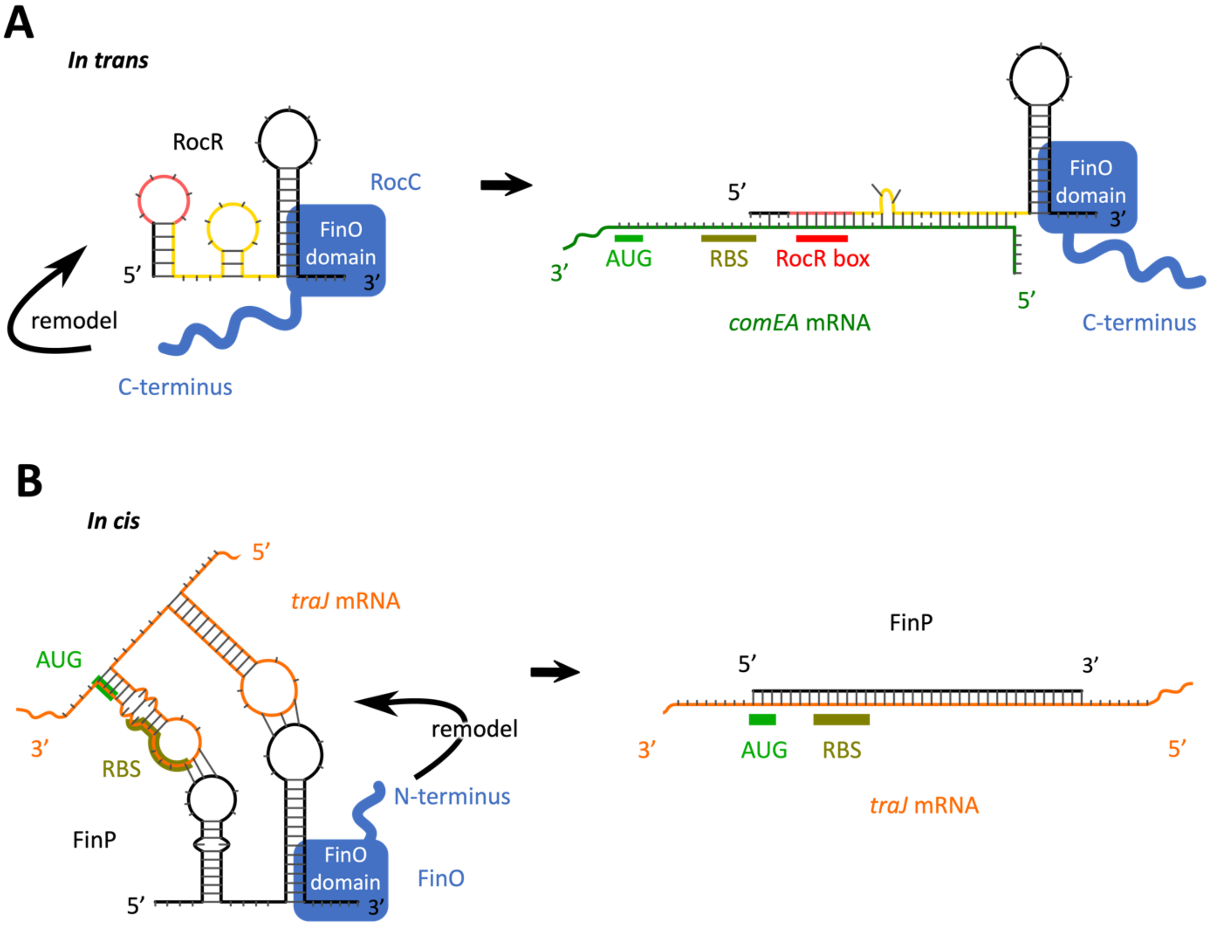


Figure 6-3. ProQ/FinO domain proteins bind transcriptional terminator structures to regulate RNA:RNA interactions.

(A) Many ProQ/FinO domain proteins, such as RocC, facilitate *in trans* RNA association between sRNAs and target mRNAs. The ProQ/FinO domain specifically binds the transcriptional terminator of the RocR sRNA, stabilizing it against degradation. Key to RNA-RNA association is the recognition of the sRNA seed sequence with its complementary region (the RocR box) in the target mRNA. The C-terminal region of RocC is required to facilitate the recognition and translational repression of target mRNAs such as *comEA*. (B) Many plasmid encoded ProQ/FinO domain proteins, such as FinO, regulate *in cis* sense-antisense RNA interactions. Similar to RocC, the ProQ/FinO domain of FinO also stabilizes the antisense RNA FinP against degradation. Initial RNA-RNA interactions are thought to involve loop:loop kissing interactions, which then proceed to duplex formation between the two RNAs, resulting in translational repression of the *traJ* mRNA target. In this case, the flexible N-terminal region of FinO is thought to be key in facilitating RNA:RNA interactions.

6.5 Concluding remarks

Specific target recognition is a crucial step to regulate gene expression and RNA interactions are especially important for post-transcriptional regulation. ProQ/FinO domain containing proteins used their ProQ/FinO domain as an RNA binding anchor to regulate sRNA-mediated gene regulation or bind directly to mRNA 3' UTRs. In this thesis, we elucidated the molecular recognition mechanism of how the RocC ProQ/FinO domain binds to intrinsic terminators. We believe that the high degree of structural similarity between ProQ/FinO domains means that the lessons learned from the RocC:RocR system will be informative for other systems. However, it remains unclear why certain proteins, such as RocC and FinO, are highly specific for their RNA targets while others, such as ProQ and NMB1681, can bind many diverse targets *in vivo*.

High selectivity is probably conferred for its irreplaceable role in physiological conditions. ProQ bound to FinP with similar binding affinity as FinO and performed chaperone activity *in vitro*, suggesting ProQ could substitute for FinO *in vivo* (Chaulk et al., 2011). *S. enterica* serovar Typhimurium is an organism that carries three ProQ/FinO proteins, ProQ, FinO, and FopA, that this speculation could be tested. RIP-seq of ProQ revealed that there are no significant ($p\text{-value} < 0.05$) reads of FinP or XepA (FinO targets), despite the fact that ProQ is more abundant than FinO (El Mouali et al., 2021). Furthermore, a ProQ null mutation does not affect FinP half-life and stability, suggesting FinO performs a distinct role, which could not be replaced by ProQ. I believe future research will elucidate the reason why ProQ and FinO have such distinct selectivity with a conserved ProQ/FinO domain structure with biological relevance.

Bibliography

Achtman, M. (1973). Transfer-positive J-independent revertants of the F factor in Escherichia coli K 12. *Genetical Research*, *21*(1), 67–77. https://doi.org/10.1017/S0016672300013227

Adams, P. D., Afonine, P. v., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, *66*(2), 213–221. https://doi.org/10.1107/S0907444909052925

Agirre, X., Román-Gómez, J., Jiménez-Velasco, A., Garate, L., Montiel-Duarte, C., Navarro, G., Vázquez, I., Zalacain, M., Calasanz, M. J., Heiniger, A., Torres, A., Minna, J. D., & Prósper, F. (2006). ASPP1, a common activator of TP53, is inactivated by aberrant methylation of its promoter in acute lymphoblastic leukemia. *Oncogene*, *25*(13), 1862–1870. https://doi.org/10.1038/sj.onc.1209236

Ahn, J., Byeon, I. J. L., Byeon, C. H., & Gronenborn, A. M. (2009). Insight into the structural basis of pro- and antiapoptotic p53 modulation by ASPP proteins. *Journal of Biological Chemistry*, *284*(20), 13812–13822. https://doi.org/10.1074/jbc.M808821200

Altegoera, F., Rensingb, S. A., & Bangea, G. (2016). Structural basis for the CsrA-dependent modulation of translation initiation by an ancient regulatory protein. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(36), 10168–10173. https://doi.org/10.1073/pnas.1602425113

Ao, Y., Rohde, L. H., & Naumovski, L. (2001). p53-interacting protein 53BP2 inhibits clonogenic survival and sensitizes cells to doxorubicin but not paclitaxel-induced apoptosis. *Oncogene*, *20*(21), 2720–2725. https://doi.org/10.1038/sj.onc.1204352

Arthur, D. C., Edwards, R. A., Tsutakawa, S., Tainer, J. A., Frost, L. S., & Glover, J. N. M. (2011). Mapping interactions between the RNA chaperone FinO and its RNA targets. *Nucleic Acids Research*, *39*(10), 4450–4463. https://doi.org/10.1093/nar/gkr025

Arthur, D. C., Ghetu, A. F., Gubbins, M. J., Edwards, R. A., Frost, L. S., & Glover, J. N. M. (2003). FinO is an RNA chaperone that facilitates sense-antisense RNA interactions. *EMBO Journal*, *22*(23), 6346–6355. https://doi.org/10.1093/emboj/cdg607

Attaiech, L., Boughammoura, A., Brochier-Armanet, C., Allatif, O., Peillard-Fiorente, F., Edwards, R. A., Omar, A. R., MacMillan, A. M., Glover, M., & Charpentier, X. (2016). Silencing of natural transformation by an RNA chaperone and a multitarget small RNA. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 8813–8818. https://doi.org/10.1073/pnas.1601626113

Attaiech, L., & Charpentier, X. (2016). Silently transformable: the many ways bacteria conceal their built-in capacity of genetic exchange. *Current Genetics*, 1–5. https://doi.org/10.1007/s00294-016-0663-6

Attaiech, L., Glover, J. N. M., & Charpentier, X. (2017). RNA Chaperones Step Out of Hfq's Shadow. *Trends in Microbiology*, *25*(4), 247–249. https://doi.org/10.1016/j.tim.2017.01.006

Aylon, Y., Ofir-Rosenfeld, Y., Yabuta, N., Lapi, E., Nojima, H., Lu, X., & Oren, M. (2010). The Lats2 tumor suppressor augments p53-mediated apoptosis by promoting the nuclear proapoptotic function of ASPP1. *Genes and Development*, *24*(21), 2420–2429. https://doi.org/10.1101/gad.1954410

Bailo, N., Kanaan, H., Kay, E., Charpentier, X., Doublet, P., & Gilbert, C. (2019). Scar-Free Genome Editing in Legionella pneumophila. *Methods in Molecular Biology (Clifton, N.J.)*, *1921*, 93–105. https://doi.org/10.1007/978-1-4939-9048-1_6

Baker, C. S., Eöry, L. A., Yakhnin, H., Mercante, J., Romeo, T., & Babitzke, P. (2007). CsrA inhibits translation initiation of Escherichia coli hfq by binding to a single site overlapping the Shine-Dalgarno sequence. *Journal of Bacteriology*, *189*(15), 5472–5481. https://doi.org/10.1128/JB.00529-07

Ban, N., Nissen, P., Hansen, J., Moore, P. B., & Steitz, T. A. (2000). The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science*, *289*(5481), 905–920. https://www.science.org

Bastet, L., Chauvier, A., Singh, N., Lussier, A., Lamontagne, A. M., Prévost, K., Massé, E., Wade, J. T., & Lafontaine, D. A. (2017). Translational control and Rho-dependent transcription termination are intimately linked in riboswitch regulation. *Nucleic Acids Research*, *45*(12), 7474–7486. https://doi.org/10.1093/nar/gkx434

Bauriedl, S., Gerovac, M., Heidrich, N., Bischler, T., Barquist, L., Vogel, J., & Schoen, C. (2020). The minimal meningococcal ProQ protein has an intrinsic capacity for structure-based global RNA recognition. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-16650-6

Bayfield, M. A., Vinayak, J., Kerkhofs, K., & Mansouri-Noori, F. (2021). La proteins couple use of sequence-specific and non-specific binding modes to engage RNA substrates. In *RNA Biology* (Vol. 18, Issue 2, pp. 168–177). Bellwether Publishing, Ltd. https://doi.org/10.1080/15476286.2019.1582955

Beisel, C. L., & Storz, G. (2010). Base pairing small RNAs and their roles in global regulatory networks. *FEMS Microbiology Reviews*, *34*(5), 866–882. https://doi.org/10.1111/j.1574-6976.2010.00241.x

Bergamaschi, D., Samuels, Y., O’Neil, N. J., Trigiante, G., Crook, T., Hsieh, J. K., O’Connor, D. J., Zhong, S., Campargue, I., Tomlinson, M. L., Kuwabara, P. E., & Lu, X. (2003). iASPP oncoprotein is a key inhibitor of p53 conserved from worm to human. *Nature Genetics*, *33*(2), 162–167. https://doi.org/10.1038/ng1070

Bergamaschi, D., Samuels, Y., Sullivan, A., Zvelebil, M., Breyssens, H., Bisso, A., Del Sal, G., Syed, N., Smith, P., Gasco, M., Crook, T., & Lu, X. (2006). iASPP preferentially binds p53 proline-rich region and modulates apoptotic function of codon 72-polymorphic p53. *Nature Genetics*, *38*(10), 1133–1141. https://doi.org/10.1038/ng1879

Berndt, N., Dohadwala, M., & Liu, C. W. Y. (1997). Constitutively active protein phosphatase 1α causes Rb-dependent G1 arrest in human cancer cells. *Current Biology*, *7*(6), 375–386. https://doi.org/10.1016/s0960-9822(06)00185-0

Bertran, M. T., Mouilleron, S., Zhou, Y., Bajaj, R., Uliana, F., Kumar, G. S., van Drogen, A., Lee, R., Banerjee, J. J., Hauri, S., O’Reilly, N., Gstaiger, M., Page, R., Peti, W., & Tapon, N. (2019). ASPP proteins discriminate between PP1 catalytic subunits through their SH3 domain and the PP1 C-tail. *Nature Communications*, *10*(1), 1–19. https://doi.org/10.1038/s41467-019-08686-0

Bieging, K. T., Mello, S. S., & Attardi, L. D. (2014). Unravelling mechanisms of p53-mediated tumour suppression. *Nature Reviews Cancer*, *14*(5), 359–370. https://doi.org/10.1038/nrc3711

Biesen, T. van, & Frost, L. S. (1994). The FinO protein of IncF plasmids binds FinP antisense RNA and its target, traJ mRNA, and promotes duplex formation. *Molecular Microbiology*, *14*(3), 427–436. https://doi.org/10.1111/j.1365-2958.1994.tb02177.x

Bollen, M., Peti, W., Ragusa, M. J., & Beullens, M. (2010). The extended PP1 toolkit: Designed to create specificity. *Trends in Biochemical Sciences*, *35*(8), 450–458. https://doi.org/10.1016/j.tibs.2010.03.002

Bolotin, A., Quinquis, B., Sorokin, A., & Dusko Ehrlich, S. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, *151*(8), 2551–2561. https://doi.org/10.1099/mic.0.28048-0

Borgstahl, G. E. O. (2007). How to use dynamic light scattering to improve the likelihood of growing macromolecular crystals. *Methods in Molecular Biology*, *363*, 109–129. https://doi.org/10.1385/1-59745-209-2:109

Bottaro, S., & Lindorff-Larsen, K. (2017). Mapping the Universe of RNA Tetraloop Folds. *Biophysical Journal*, *113*(2), 257–267. https://doi.org/10.1016/j.bpj.2017.06.011

Brautigan, D. L., & Shenolikar, S. (2018). Protein Serine/Threonine Phosphatases: Keys to Unlocking Regulators and Substrates. *Annual Review of Biochemistry*, *87*, 921–964. https://doi.org/10.1146/annurev-biochem-062917-012332

Brockdorff, N., Ashworth, A., Kay, G. F., Mccabe, V. M., Norris, D. P., Cooper, P. J., Swift, S., & Rastan', S. (1992). The Product of the Mouse Xist Gene Is a 15 kb Inactive X-Specific Transcript Containing No Consewed ORF and Located in the Nucleus. In *Cell* (Vol. 71).

Buchan, D. W. A., Minneci, F., Nugent, T. C. O., Bryson, K., & Jones, D. T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Research*, *41*(Web Server issue), 349–357. https://doi.org/10.1093/nar/gkt381

Bunkóczi, G., & Read, R. J. (2011). Improvement of molecular-replacement models with Sculptor. *Acta Crystallographica Section D: Biological Crystallography*, *67*(4), 303–312. https://doi.org/10.1107/S0907444910051218

Cao, L., Huang, Q., He, J., Lu, J., & Xiong, Y. (2013). Elevated expression of iASPP correlates with poor prognosis and chemoresistance/radioresistance in FIGO Ib1-IIa squamous cell cervical cancer. *Cell and Tissue Research*, *352*(2), 361–369. https://doi.org/10.1007/s00441-013-1569-y

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., Doudna, J. A., Cate, J. H., Zhou, K., Doudna are, J. A., Gooding, A. R., Podell, E., Golden, B. L., Kundrot, C. E., & Cech are, T. R. (1996). Crystal Structure of a Group I Ribozyme Domain: Principles of RNA Packing. *Science*, *273*(5282), 1678–11685. https://www.science.org

Chao, Y., Papenfort, K., Reinhardt, R., Sharma, C. M., & Vogel, J. (2012). An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *The EMBO Journal*, *31*(20), 4005–4019. https://doi.org/10.1038/emboj.2012.229

Chaulk, S., Lu, J., Tan, K., Arthur, D. C., Edwards, R. a, Frost, L. S., Joachimiak, A., & Glover, J. N. M. (2010). N. meningitidis 1681 is a member of the FinO family of RNA chaperones. *RNA Biology*, *7*(6), 812–819. https://doi.org/10.4161/rna.7.6.13688

Chaulk, S., Smith-Frieday, M. N., Arthur, D. C., Culham, D. E., Edwards, R. A., Soo, P., Frost, L. S., Keates, R. A. B., Glover, J. N. M., & Wood, J. M. (2011). ProQ is an RNA chaperone that controls ProP levels in escherichia coli. *Biochemistry*, *50*(15), 3095–3106. https://doi.org/10.1021/bi101683a

Chayen, N. E., & Saridakis, E. (2008). Protein crystallization: From purified protein to diffraction-quality crystal. *Nature Methods*, *5*(2), 147–153. https://doi.org/10.1038/nmeth.f.203

Choy, M. S., Hieke, M., Kumar, G. S., Lewis, G. R., Gonzalez-DeWhitt, K. R., Kessler, R. P., Stein, B. J., Hessenberger, M., Nairn, A. C., Peti, W., & Page, R. (2014). Understanding the antagonism of retinoblastoma protein dephosphorylation by PNUTS provides insights into the PP1 regulatory code. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(11), 4097–4102. https://doi.org/10.1073/pnas.1317395111

Choy, M. S., Page, R., & Peti, W. (2012). Regulation of protein phosphatase 1 by intrinsically disordered proteins. *Biochemical Society Transactions*, *40*(5), 969–974. https://doi.org/10.1042/BST20120094

Chris Oubridge, Nobutoshi Ito, Philip R. Evans, C.-Hiang Teo, & Kiyoshi Nagai. (1994). Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, *372*(6505), 432–438.

Christopher, G. K., Phipps, A. G., & Gray, R. J. (1998). Temperature-dependent solubility of selected proteins. *Journal of Crystal Growth*, *191*(4), 820–826. https://doi.org/10.1016/S0022-0248(98)00355-8

Chruszcz, M., Potrzebowski, W., Zimmerman, M. D., Grabowski, M., Zheng, H., Lasota, P., & Minor, W. (2008). Analysis of solvent content and oligomeric states in protein crystals-does symmetry matter? *Protein Science*, *17*(4), 623–632. https://doi.org/10.1110/ps.073360508

Cléry, A., & Allain, F. H. -T. (2012). From Structure to Function of RNA Binding Domains. In *In RNA Binding Proteins, Z.J. Lorkovic, ed. (Landes Bioscience)*. https://doi.org/10.1201/9781498713368-15

Cléry, A., Blatter, M., & Allain, F. H. T. (2008). RNA recognition motifs: boring? Not quite. In *Current Opinion in Structural Biology* (Vol. 18, Issue 3, pp. 290–298). https://doi.org/10.1016/j.sbi.2008.04.002

Cui, M., Wang, T., Xu, J., Ke, Y., Du, X., Yuan, X., Wang, Z., Gong, C., Zhuang, Y., Lei, S., Su, X., Wang, X., Huang, L., Zhong, Z., Peng, G., Yuan, J., Chen, Z., & Wang, Y. (2013). Impact of Hfq on global gene expression and intracellular survival in Brucella melitensis. *PloS One*, *8*(8). https://doi.org/10.1371/journal.pone.0071933

Dancheck, B., Ragusa, M. J., Allaire, M., Nairn, A. C., Page, R., & Peti, W. (2011). Molecular investigations of the structure and function of the protein phosphatase 1-spinophilin-inhibitor 2 heterotrimeric complex. *Biochemistry*, *50*(7), 1238–1246. https://doi.org/10.1021/bi101774g

Daniel Aceytuno, R., Piett, C. G., Havali-Shahriari, Z., Edwards, R. A., Rey, M., Ye, R., Javed, F., Fang, S., Mani, R., Weinfeld, M., Hammel, M., Tainer, J. A., Schriemer, D. C., Lees-Miller, S. P., & Mark Glover, J. N. (2017). Structural and functional characterization of the

PNKP-XRCC4-LigIV DNA repair complex. *Nucleic Acids Research*, *45*(10), 6238–6251. https://doi.org/10.1093/nar/gkx275

De Lay, N., & Gottesman, S. (2012). A complex network of small non-coding RNAs regulate motility in Escherichia coli. *Molecular Microbiology*, *86*(3), 524–538. https://doi.org/10.1111/j.1365-2958.2012.08209.x

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., & Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, *471*(7340), 602–607. https://doi.org/10.1038/nature09886

Dimastrogiovanni, D., Fröhlich, K. S., Bandyra, K. J., Bruce, H. A., Hohensee, S., Vogel, J., & Luisi, B. F. (2014). Recognition of the small regulatory RNA RydC by the bacterial Hfq protein. *ELife*, *3*, 1–19. https://doi.org/10.7554/eLife.05375

Dohadwala, M., Da Cruz Silva, E. E. F., Hall, F. L., Williams, R. T., Carbonaro-Hall, D. A., Nairn, A. C., Greengard, P., & Berndt, N. (1994). Phosphorylation and inactivation of protein phosphatase 1 by cyclin- dependent kinases. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(14), 6408–6412. https://doi.org/10.1073/pnas.91.14.6408

Dong, G., Chakshusmathi, G., Wolin, S. L., & Reinisch, K. M. (2004). Structure of the La motif: A winged helix domain mediates RNA binding via a conserved aromatic patch. *EMBO Journal*, *23*(5), 1000–1007. https://doi.org/10.1038/sj.emboj.7600115

Dubey, A. K., Baker, C. S., Romeo, T., & Babitzke, P. (2005). RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. *Rna*, *11*(10), 1579–1587. https://doi.org/10.1261/rna.2990205

Easton, L. E., Shibata, Y., & Lukavsky, P. J. (2010). Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. *Rna*, *16*(3), 647–653. https://doi.org/10.1261/rna.1862210

Edwards, A. N., Patterson-Fortin, L. M., Vakulskas, C. A., Mercante, J. W., Potrykus, K., Vinella, D., Camacho, M. I., Fields, J. A., Thompson, S. A., Georgellis, D., Cashel, M., Babitzke, P., & Romeo, T. (2011). Circuitry linking the Csr and stringent response global regulatory systems. *Molecular Microbiology*, *80*(6), 1561–1580. https://doi.org/10.1111/j.1365-2958.2011.07663.x

el Mouali, Y., Gerovac, M., Mineikaite, R., & Vogel, J. (2021). In vivo targets of Salmonella FinO include a FinP-like small RNA controlling copy number of a cohabitating plasmid. *Nucleic Acids Research*, *49*(9), 5319–5335. https://doi.org/10.1093/nar/gkab281

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, *66*(4), 486–501. https://doi.org/10.1107/S0907444910007493

Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N., & Nordlund, P. (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry*, *357*(2), 289–298. https://doi.org/10.1016/j.ab.2006.07.027

Espanel, X., & Sudol, M. (2001). Yes-associated Protein and p53-binding Protein-2 Interact through Their WW and SH3 Domains. *Journal of Biological Chemistry*, *276*(17), 14514–14523. https://doi.org/10.1074/jbc.M008568200

Feng, S., Chen, J. K., Yu, H., Simon, J. A., & Schreiber, S. L. (1994). Two binding orientations for peptides to the Src SH3 domain: Development of a general model for SH3-ligand interactions. *Science*, *266*(5188), 1241–1247. https://doi.org/10.1126/science.7526465

Finlay, B. B., Frost, L. S., Paranchych, W., & Willetts, N. S. (1986). Nucleotide sequences of five IncF plasmid finP alleles. *Journal of Bacteriology*, *167*(2), 754–757.

Finnegan, D. J., & Willetts, N. S. (1971). Two classes of F lac mutants insensitive to transfer inhibition by an F-like R factor. *MGG Molecular & General Genetics*, *111*(3), 256–264. https://doi.org/10.1007/BF00433110

Franze De Fernandez, M. T., Eoyang, L., & August, J. T. (1968). Factor fraction required for the synthesis of bacteriophage Qβ-RNA. *Nature*, *219*(5154), 588–590. https://doi.org/10.1038/219588a0

Frost, L. S., Ippen-Ihler, K., & Skurray, R. A. (1994). Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiological Reviews*, *58*(2), 162–210. https://doi.org/10.1128/mr.58.2.162-210.1994

Gan, J., Tropea, J. E., Austin, B. P., Court, D. L., Waugh, D. S., & Ji, X. (2006). Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. *Cell*, *124*(2), 355–366. https://doi.org/10.1016/j.cell.2005.11.034

Garst, A. D., Edwards, A. L., & Batey, R. T. (2011). Riboswitches: Structures and mechanisms. In *Cold Spring Harbor Perspectives in Biology* (Vol. 3, Issue 6, pp. 1–13). Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/cshperspect.a003533

Geissmann, T. A., & Touati, D. (2004). Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. *The EMBO Journal*, *23*(2), 396–405. https://doi.org/10.1038/sj.emboj.7600058

Ghetu, A. F., Arthur, D. C., Kerppola, T. K., & Glover, J. N. M. (2002). Probing FinO-FinP RNA interactions by site-directed protein-RNA crosslinking and gelFRET. *RNA (New York, N.Y.)*, *8*(6), 816–823. https://doi.org/10.1017/S1355838202026730

Ghetu, A. F., Gubbins, M. J., Frost, L. S., & Glover, J. N. M. (2000). Crystal structure of the bacterial conjugation repressor FinO. *Nature Structural Biology*, *7*(7), 565–569. https://doi.org/10.1038/76790

Ghetu, A. F., Gubbins, M. J., Oikawa, K., Kay, C. M., Frost, L. S., & Glover, J. N. M. (1999). The FinO repressor of bacterial conjugation contains two RNA binding regions. *Biochemistry*, *38*(42), 14036–14044. https://doi.org/10.1021/bi9911482

Godin-Heymann, N., Wang, Y., Slee, E., & Lu, X. (2013). Phosphorylation of ASPP2 by RAS/MAPK pathway is critical for its full pro-apoptotic function. *PLoS ONE*, *8*(12), 1–10. https://doi.org/10.1371/journal.pone.0082022

Golden, B. L. (2007). *Preparation and Crystalli z ation of RNA*. *363*.

Gonzalez, G. M., Hardwick, S. W., Maslen, S. L., Mark Skehel, J., Holmqvist, E., Vogel, J., Bateman, A., Luisi, B. F., & William Broadhurst, R. (2017). Structure of the Escherichia coli ProQ RNA-binding protein. *RNA*, *23*(5), 696–711. https://doi.org/10.1261/rna.060343.116

Gorina, S., & Pavletich, N. P. (1996). Structure of the p53 tumor suppressor bound to the ankyrin and SH3 domains of 53BP2. *Science*, *274*(5289), 1001–1005. https://doi.org/10.1126/science.274.5289.1001

Gorrec, F. (2009). The MORPHEUS protein crystallization screen. *Journal of Applied Crystallography*, *42*(6), 1035–1042. https://doi.org/10.1107/S0021889809042022

Gottesman, S., & Storz, G. (2011). Bacterial small RNA regulators: Versatile roles and rapidly evolving variations. *Cold Spring Harbor Perspectives in Biology*, *3*(12), 1–16. https://doi.org/10.1101/cshperspect.a003798

Grassie, M. E., Moffat, L. D., Walsh, M. P., & MacDonald, J. A. (2011). The myosin phosphatase targeting protein (MYPT) family: A regulated mechanism for achieving substrate specificity of the catalytic subunit of protein phosphatase type 1δ. *Archives of Biochemistry and Biophysics*, *510*(2), 147–159. https://doi.org/10.1016/j.abb.2011.01.018

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R., & Bateman, A. (2005). Rfam: Annotating non-coding RNAs in complete genomes. *Nucleic Acids Research*, *33*(DATABASE ISS.), 121–124. https://doi.org/10.1093/nar/gki081

Gubbins, M. J., Arthur, D. C., Ghetu, A. F., Glover, J. N. M., & Frost, L. S. (2003). Characterizing the structural features of RNA/RNA interactions of the F-plasmid FinOP fertility inhibition system. *Journal of Biological Chemistry*, *278*(30), 27663–27671. https://doi.org/10.1074/jbc.M303186200

Gutiérrez, P., Li, Y., Osborne, M. J., Pomerantseva, E., Liu, Q., & Gehring, K. (2005). Solution structure of the carbon storage regulator protein CsrA from Escherichia coli. *Journal of Bacteriology*, *187*(10), 3496–3501. https://doi.org/10.1128/JB.187.10.3496-3501.2005

Hämmerle, H., Amman, F., Večerek, B., Stülke, J., Hofacker, I., & Bläsi, U. (2014). Impact of Hfq on the Bacillus subtilis transcriptome. *PLoS ONE*, *9*(6). https://doi.org/10.1371/journal.pone.0098661

Härtig, E., & Jahn, D. (2012). Regulation of the Anaerobic Metabolism in Bacillus subtilis. *Advances in Microbial Physiology*, *61*, 195–216. https://doi.org/10.1016/B978-0-12-394423-8.00005-6

Heeb, S., Kuehne, S. A., Bycroft, M., Crivii, S., Allen, M. D., Haas, D., Cámara, M., & Williams, P. (2006). Functional analysis of the post-transcriptional regulator RsmA reveals a novel RNA-binding site. *Journal of Molecular Biology*, *355*(5), 1026–1036. https://doi.org/10.1016/j.jmb.2005.11.045

Helps, N. R., Barker, H. M., Elledge, S. J., & Cohen, P. T. W. (1995). Protein phosphatase 1 interacts with p53BP2, a protein which binds to the tumour suppressor p53. *FEBS Letters*, *377*(3), 295–300. https://doi.org/10.1016/0014-5793(95)01347-4

Herschlag, D. (1995). RNA chaperones and the folding problem. *Journal of Biological Chemistry*, *270*, 20871–20874. https://doi.org/10.1074/jbc.270.36.20871

Hodge, C. D., Ismail, I. H., Edwards, R. A., Hura, G. L., Xiao, A. T., Tainer, J. A., Hendzel, M. J., & Glover, J. N. M. (2016). RNF8 E3 ubiquitin ligase stimulates Ubc13 E2 conjugating activity that is essential for DNA double strand break signaling and BRCA1 tumor suppressor recruitment. *Journal of Biological Chemistry*, *291*(18), 9396–9410. https://doi.org/10.1074/jbc.M116.715698

HOLLEY, R. W., APGAR, J., EVERETT, G. A., MADISON, J. T., MARQUISEE, M., MERRILL, S. H., PENSWICKAND, J. R., & ZAMIR, A. (1965). Structure of a Ribonucleic Acid. *Science*, *147*, 1462–1465.

Holmqvist, E., Li, L., Bischler, T., Barquist, L., & Vogel, J. (2018). Global Maps of ProQ Binding In Vivo Reveal Target Recognition via RNA Structure and Stability Control at mRNA 3′ Ends. *Molecular Cell*, *70*(5), 971-982.e6. https://doi.org/10.1016/j.molcel.2018.04.017

Holmqvist, E., & Vogel, J. (2018). RNA-binding proteins in bacteria. In *Nature Reviews Microbiology* (Vol. 16, Issue 10, pp. 601–615). Nature Publishing Group. https://doi.org/10.1038/s41579-018-0049-5

Holmqvist, E., Wright, P. R., Li, L., Bischler, T., Barquist, L., Reinhardt, R., Backofen, R., & Vogel, J. (2016). Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo . *The EMBO Journal*, *35*(9), 991–1011. https://doi.org/10.15252/embj.201593360

Hopkins, J. B., Gillilan, R. E., & Skou, S. (2017). BioXTAS RAW: Improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *Journal of Applied Crystallography*, *50*(5), 1545–1553. https://doi.org/10.1107/S1600576717011438

Hör, J., Gorski, S. A., & Vogel, J. (2018). Bacterial RNA Biology on a Genome Scale. *Molecular Cell*, *70*(5), 785–799. https://doi.org/10.1016/j.molcel.2017.12.023

Horstmann, N., Orans, J., Valentin-Hansen, P., Shelburne, S. A., & Brennan, R. G. (2012). Structural mechanism of Staphylococcus aureus Hfq binding to an RNA A-tract. *Nucleic Acids Research*, *40*(21), 11023–11035. https://doi.org/10.1093/nar/gks809

Hurley, T. D., Yang, J., Zhang, L., Goodwin, K. D., Zou, Q., Cortese, M., Dunker, A. K., & DePaoli-Roach, A. A. (2007). Structural basis for regulation of protein phosphatase 1 by inhibitor-2. *Journal of Biological Chemistry*, *282*(39), 28874–28883. https://doi.org/10.1074/jbc.M703472200

Imburgio, D., Rong, M., Ma, K., and McAllister, W. T. (2000). Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry*, 39, 10419–10430

Immer, C., Hacker, C., & Wöhnert, J. (2020). Solution structure and RNA-binding of a minimal ProQ-homolog from Legionella pneumophila (Lpp1663). *RNA (New York, N.Y.)*, *26*(12), 2031–2043. https://doi.org/10.1261/rna.077354.120

Inouye, M., & Delihast, N. (1988). Small RNAs in the Prokaryotes: A Growing List of Diverse Roles Minireview. *Cell*, *53*(1), 5–7.

Ishikawa, H., Otaka, H., Maki, K., Morita, T., & Aiba, H. (2012). The functional Hfq-binding module of bacterial sRNAs consists of a double or single hairpin preceded by a U-rich sequence and followed by a 3′ poly(U) tail. *Rna*, *18*(5), 1062–1074. https://doi.org/10.1261/rna.031575.111

Ishino, Y., Krupovic, M., & Forterre, P. (2018). History of CRISPR-Cas from encounter with a mysterious repeated sequence to genome editing technology. *Journal of Bacteriology*, *200*(7). https://doi.org/10.1128/JB.00580-17

Ito, M., Nakano, T., Erdodi, F., & Hartshorne, D. J. (2004). Myosin phosphatase: Structure, regulation and function. *Molecular and Cellular Biochemistry*, *259*(1–2), 197–209. https://doi.org/10.1023/B:MCBI.0000021373.14288.00

Jerome, L. J., & Frost, L. S. (1999). In vitro analysis of the interaction between the FinO protein and FinP antisense RNA of F-like conjugative plasmids. *Journal of Biological Chemistry*, *274*(15), 10356–10362. https://doi.org/10.1074/jbc.274.15.10356

Jerome, L. J., Van Biesen, T., & Frost, L. S. (1999). Degradation of FinP antisense RNA from F-like plasmids: The RNA-binding protein, FinO, protects FinP from ribonuclease E. *Journal of Molecular Biology*, *285*(4), 1457–1473. https://doi.org/10.1006/jmbi.1998.2404

Jia, Y., Peng, L., Rao, Q., Xing, H., Huai, L., Yu, P., Chen, Y., Wang, C., Wang, M., Mi, Y., & Wang, J. (2014). Oncogene iASPP enhances self-renewal of hematopoietic stem cells and facilitates their resistance to chemotherapy and irradiation. *FASEB Journal*, *28*(7), 2816–2827. https://doi.org/10.1096/fj.13-244632

Jiang, F., & Doudna, J. A. (2017). *CRISPR-Cas9 Structures and Mechanisms*. https://doi.org/10.1146/annurev-biophys

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, *337*, 816–821. https://www.science.org

Karen Montzka Wassarman, Aixia Zhang, & Gisela Storz. (1999). Small RNAs in Escherichia coli. *Trends in Microbiology*, *7*(1), 37–45.

Katsuya-gaviria, K., Paris, G., Dendooven, T., Bandyra, K. J., Paris, G., Dendooven, T., Katarzyna, J., Katsuya-gaviria, K., Paris, G., & Dendooven, T. (2022). Bacterial RNA chaperones and chaperone- like riboregulators : behind the scenes of RNA- mediated regulation of cellular metabolism RNA-mediated regulation of cellular metabolism. *RNA Biology*, *19*(1), 419–436. https://doi.org/10.1080/15476286.2022.2048565

Katz, C., Benyamini, H., Rotem, S., Lebendiker, M., Danieli, T., Iosub, A., Refaely, H., Dines, M., Bronner, V., Bravman, T., Shalev, D. E., Rüdiger, S., & Friedler, A. (2008). Molecular basis of the interaction between the antiapoptotic Bcl-2 family proteins and the proapoptotic protein ASPP2. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(34), 12277–12282. https://doi.org/10.1073/pnas.0711269105

Kavita, K., de Mets, F., & Gottesman, S. (2018). New aspects of RNA-based regulation by Hfq and its partner sRNAs. *Current Opinion in Microbiology*, *42*, 53–61. https://doi.org/10.1016/j.mib.2017.10.014

Kim, H. J., Chaulk, S., Arthur, D., Edwards, R. A., & Glover, J. N. M. (2020). Biochemical Methods for the Study of the FinO Family of Bacterial RNA Chaperones. In T. Heise (Ed.), *RNA Chaperones: Methods and Protocols* (pp. 1–18). Springer US. https://doi.org/10.1007/978-1-0716-0231-7_1

Kim, S. H., Suddath, F. L., Quigley, G. J., Sussman, J. L., Wang, A. H. J., Seeman, N. C., & Rich, A. (1974). Three-Dimensional Tertiary Structure of Yeast Phenylalanine Transfer RNA. In *Science* (Vol. 185, Issue August).

Koraimann, G., Koraimann, C., Koronakis, V., Schlager, S., & Högenauer, G. (1991). Repression and derepression of conjugation of plasmid R1 by wild-type and mutated finP antisense RNA. *Molecular Microbiology*, *5*(1), 77–87. https://doi.org/10.1111/j.1365-2958.1991.tb01828.x

Kovach, A. R., Hoff, K. E., Canty, J. T., Orans, J., & Brennan, R. G. (2014). Recognition of U-rich RNA by Hfq from the Gram-positive pathogen Listeria monocytogenes. *Rna*, *20*(10), 1548–1559. https://doi.org/10.1261/rna.044032.113

Kunte, H. J., Crane, R. A., Culham, D. E., Richmond, D., & Wood, J. M. (1999). Protein ProQ influences osmotic activation of compatible solute transporter ProP in Escherichia coli K-12. *Journal of Bacteriology*, *181*(5), 1537–1543.

Lees-Miller, S. P., Sakaguchi, K., Ullrich, S. J., Appella, E., & Anderson, C. W. (1992). Human DNA-activated protein kinase phosphorylates serines 15 and 37 in the amino-terminal transactivation domain of human p53. *Molecular and Cellular Biology*, *12*(11), 5041–5049. https://doi.org/10.1128/mcb.12.11.5041-5049.1992

Leung, A. K. W., Nagai, K., & Li, J. (2011). Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. *Nature*, *473*(7348), 536–539. https://doi.org/10.1038/nature09956

Li, D. W. C., Liu, J. P., Schmid, P. C., Schlosser, R., Feng, H., Liu, W. B., Yan, Q., Gong, L., Sun, S. M., Deng, M., & Liu, Y. (2006). Protein serine/threonine phosphatase-1 dephosphorylates p53 at Ser-15 and Ser-37 to modulate its transcriptional and apoptotic activities. *Oncogene*, *25*(21), 3006–3022. https://doi.org/10.1038/sj.onc.1209334

Li, G., Wang, R., Gao, J., Deng, K., Wei, J., & Wei, Y. (2011). RNA interference-mediated silencing of iASPP induces cell proliferation inhibition and G0/G1 cell cycle arrest in U251 human glioblastoma cells. *Molecular and Cellular Biochemistry*, *350*(1–2), 193–200. https://doi.org/10.1007/s11010-010-0698-9

Li, S., Shi, G., Yuan, H., Zhou, T., Zhang, Q., Zhu, H., & Wang, X. (2012). Abnormal expression pattern of the ASPP family of proteins in human non-small cell lung cancer and regulatory functions on apoptosis through p53 by iASPP. *Oncology Reports*, *28*(1), 133–140. https://doi.org/10.3892/or.2012.1778

Liang, J. C., Bloom, R. J., & Smolke, C. D. (2011). Engineering Biological Systems with Synthetic RNA Molecules. *Molecular Cell*, *43*(6), 915–926. https://doi.org/10.1016/j.molcel.2011.08.023

Lim, W. A., Richards, F. M., & Fox, R. (1994). Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature*, *372*, 375–379.

LIN, B. L., XIE, D. Y., XIE, S. B., XIE, J. Q., ZHANG, X. H., ZHANG, Y. F., & GAO, Z. L. (2011). Down-regulation of iASPP in human hepatocellular carcinoma cells inhibits cell proliferation and tumor growth. *Neoplasma*, *58*(3), 205–210. https://doi.org/10.4149/neo

Link, T. M., Valentin-Hansen, P., & Brennana, R. G. (2009). Structure of Escherichia coli Hfq bound to polyriboadenylate RNA. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(46), 19292–19297. https://doi.org/10.1073/pnas.0908744106

Liu, H., Wang, M., Diao, S., Rao, Q., Zhang, X., Xing, H., & Wang, J. (2009). siRNA-mediated down-regulation of iASPP promotes apoptosis induced by etoposide and daunorubicin in leukemia cells expressing wild-type p53. *Leukemia Research*, *33*(9), 1243–1248. https://doi.org/10.1016/j.leukres.2009.02.016

Liu, W. K., Jiang, X. Y., Ren, J. K., & Zhang, Z. X. (2010). Expression pattern of the ASPP family members in endometrial endometrioid adenocarcinoma. *Onkologie*, *33*(10), 500–503. https://doi.org/10.1159/000319692

Liu, Z. J., Lu, X., Zhang, Y., Zhong, S., Gu, S. Z., Zhang, X. B., Yang, X., & Xin, H. M. (2005). Downregulated mRNA expression of ASPP and the hypermethylation of the 5′-untranslated region in cancer cell lines retaining wild-type p53. *FEBS Letters*, *579*(7), 1587–1590. https://doi.org/10.1016/j.febslet.2005.01.069

Liu, Z. J., Zhang, Y., Zhang, X. B., & Yang, X. (2004). Abnormal mRNA expression of ASPP members in leukemia cell lines. *Leukemia*, *18*(4), 880. https://doi.org/10.1038/sj.leu.2403300

Llanos, S., Royer, C., Lu, M., Bergamaschi, D., Lee, W. H., & Lu, X. (2011). Inhibitory member of the apoptosis-stimulating proteins of the p53 family (iASPP) interacts with protein phosphatase 1 via a noncanonical binding motif. *Journal of Biological Chemistry*, *286*(50), 43039–43044. https://doi.org/10.1074/jbc.M111.270751

Lopez, D., Vlamakis, H., & Kolter, R. (2009). Generation of multiple cell types in Bacillus subtilis. *FEMS Microbiology Reviews*, *33*(1), 152–163. https://doi.org/10.1111/j.1574-6976.2008.00148.x

Lorenz, C., Gesell, T., Zimmermann, B., Schoeberl, U., Bilusic, I., Rajkowitsch, L., Waldsich, C., von Haeseler, A., & Schroeder, R. (2010). Genomic SELEX for Hfq-binding RNAs identifies genomic aptamers predominantly in antisense transcripts. *Nucleic Acids Research*, *38*(11), 3794–3808. https://doi.org/10.1093/nar/gkq032

Lossos, I. S., Natkunam, Y., Levy, R., & Lopez, C. D. (2002). Apoptosis stimulating protein of p53 (ASPP2) expression differs in diffuse large B-cel and follicular center lymphoma: Correlation with clinical outcome. *Leukemia and Lymphoma*, *43*(12), 2309–2317. https://doi.org/10.1080/1042819021000040017

Lu, B., Guo, H., Zhao, J., Wang, C., Wu, G., Pang, M., Tong, X., Bu, F., Liang, A., Hou, S., Fan, X., Dai, J., Wang, H., & Guo, Y. (2010). Increased expression of iASPP, regulated by hepatitis B virus X protein-mediated NF-κB activation, in hepatocellular carcinoma. *Gastroenterology*, *139*(6), 2183–2194. https://doi.org/10.1053/j.gastro.2010.06.049

Lu, M., Breyssens, H., Salter, V., Zhong, S., Hu, Y., Baer, C., Ratnayaka, I., Sullivan, A., Brown, N. R., Endicott, J., Knapp, S., Kessler, B. M., Middleton, M. R., Siebold, C., Jones, E. Y., Sviderskaya, E. V., Cebon, J., John, T., Caballero, O. L., … Lu, X. (2013). Restoring p53 Function in Human Melanoma Cells by Inhibiting MDM2 and Cyclin B1/CDK1-Phosphorylated Nuclear iASPP. *Cancer Cell*, *23*(5), 618–633. https://doi.org/10.1016/j.ccr.2013.03.013

Luscombe, N. M., Laskowski, R. A., & Thornton, J. M. (2001). Amino acid-base interactions: A three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Research*, *29*(13), 2860–2874. https://doi.org/10.1093/nar/29.13.2860

Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C., & Breaker, R. R. (2003). Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. *Cell*, *113*(5), 577–586. https://doi.org/10.1016/S0092-8674(03)00391-X

Maraia, R. J., & Bayfield, M. A. (2006). The La Protein-RNA complex surfaces. In *Molecular Cell* (Vol. 21, Issue 2, pp. 149–152). https://doi.org/10.1016/j.molcel.2006.01.004

Maraia, R. J., & Intine, R. V. A. (2001). Recognition of Nascent RNA by the Human La Antigen: Conserved and Divergent Features of Structure and Function. *Molecular and Cellular Biology*, *21*(2), 367–379. https://doi.org/10.1128/mcb.21.2.367-379.2001

Maris, C., Dominguez, C., & Allain, F. H. T. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. In *FEBS Journal* (Vol. 272, Issue 9, pp. 2118–2131). https://doi.org/10.1111/j.1742-4658.2005.04653.x

Mark Glover, J. N., Chaulk, S. G., Edwards, R. A., Arthur, D., Lu, J., & Frost, L. S. (2015). The FinO family of bacterial RNA chaperones. *Plasmid*, *78*, 79–87. https://doi.org/10.1016/j.plasmid.2014.07.003

Martins, P. M., Rocha, F., & Damas, A. M. (2008). Understanding water equilibration fundamentals as a step for rational protein crystallization. *PLoS ONE*, *3*(4). https://doi.org/10.1371/journal.pone.0001998

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, *40*(4), 658–674. https://doi.org/10.1107/S0021889807021206

McPherson, A., & Gavira, J. A. (2014). Introduction to protein crystallization. *Acta Crystallographica Section F:Structural Biology Communications*, *70*(1), 2–20. https://doi.org/10.1107/S2053230X13033141

Mercante, J., Edwards, A. N., Dubey, A. K., Babitzke, P., & Romeo, T. (2009). Molecular Geometry of CsrA (RsmA) Binding to RNA and Its Implications for Regulated Expression. *Journal of Molecular Biology*, *392*(2), 511–528. https://doi.org/10.1016/j.jmb.2009.07.034

Meynell, E., & Datta, N. (1967). Mutant drug resistant factors of high transmissibility. *Nature*, *214*(5091), 885–887. https://doi.org/10.1038/214885a0

Mironov, A. S., Gusarov, I., Rafikov, R., Lopez, L. E., Shatalin, K., Kreneva, R. A., Perumov, D. A., & Nudler, E. (2002). Sensing small molecules by nascent RNA: A mechanism to control transcription in bacteria. *Cell*, *111*(5), 747–756. https://doi.org/10.1016/S0092-8674(02)01134-0

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution*, *60*(2), 174–182. https://doi.org/10.1007/s00239-004-0046-3

Mori, S., Ito, G., Usami, N., Yoshioka, H., Ueda, Y., Kodama, Y., Takahashi, M., Fong, K. M., Shimokata, K., & Sekido, Y. (2004). p53 Apoptotic Pathway Molecules Are Frequently and Simultaneously Altered in Nonsmall Cell Lung Carcinoma. *Cancer*, *100*(8), 1673–1682. https://doi.org/10.1002/cncr.20164

Mori, T., Okamoto, H., Takahashi, N., Ueda, R., & Okamoto, T. (2000). Aberrant overexpression of 53BP2 mRNA in lung cancer cell lines. *FEBS Letters*, *465*(2–3), 124–128. https://doi.org/10.1016/S0014-5793(99)01726-3

Morris, K. v., & Mattick, J. S. (2014). The rise of regulatory RNA. In *Nature Reviews Genetics* (Vol. 15, Issue 6, pp. 423–437). Nature Publishing Group. https://doi.org/10.1038/nrg3722

Mui, M. Z., Zhou, Y., Blanchette, P., Chughtai, N., Knight, J. F., Gruosso, T., Papadakis, A. I., Huang, S., Park, M., Gingras, A.-C., & Branton, P. E. (2015). The Human Adenovirus Type 5 E4orf4 Protein Targets Two Phosphatase Regulators of the Hippo Signaling Pathway. *Journal of Virology*, *89*(17), 8855–8870. https://doi.org/10.1128/jvi.03710-14

Mullineaux, P., & Willetts, N. (1985). Promoters in the transfer region of plasmid F. *Plasmids in Bacteria*, *In: Helinski DR, Cohen SN, Clewelt DB, Jackson DA, Hollaender A (eds) Plasmids in bacteria. Plenum Publishing Corp, New York*, 605–614.

Mura, C., Randolph, P. S., Patterson, J., & Cozen, A. E. (2013). Archaeal and eukaryotic homologs of Hfq: A structural and evolutionary perspective on Sm function. *RNA Biology*, *10*(4), 636–651. https://doi.org/10.4161/rna.24538

Murina, V., Lekontseva, N., & Nikulin, A. (2013). Hfq binds ribonucleotides in three different RNA-binding sites. *Acta Crystallographica Section D: Biological Crystallography*, *69*(8), 1504–1513. https://doi.org/10.1107/S090744491301010X

Muto, Y., & Yokoyama, S. (2012). Structural insight into RNA recognition motifs: Versatile molecular Lego building blocks for biological systems. In *Wiley Interdisciplinary Reviews: RNA* (Vol. 3, Issue 2, pp. 229–246). https://doi.org/10.1002/wrna.1107

Narberhaus, F., & Vogel, J. (2009). Regulatory RNAs in prokaryotes: Here, there and everywhere: Micromeeting report. *Molecular Microbiology*, *74*(2), 261–269. https://doi.org/10.1111/j.1365-2958.2009.06869.x

Naumovski, L., & Cleary, M. L. (1996). The p53-binding protein 53BP2 also interacts with Bc12 and impedes cell cycle progression at G2/M. *Molecular and Cellular Biology*, *16*(7), 3884–3892. https://doi.org/10.1128/mcb.16.7.3884

Nowak, R. (1994). Mining treasures from "junk DNA." *Science*, *263*, 608–610. http://science.sciencemag.org/

Ohndorf, U. M., Steegborn, C., Knijff, R., & Sondermann, P. (2001). Contributions of the Individual Domains in Human la Protein to Its RNA 3′-End Binding Activity. *Journal of Biological Chemistry*, *276*(29), 27188–27196. https://doi.org/10.1074/jbc.M102891200

Olejniczak, M., & Storz, G. (2017). ProQ/FinO-domain proteins: Another ubiquitous family of RNA matchmakers? *Molecular Microbiology*, *104*, 905–915. https://doi.org/10.1111/mmi.13679

Ontiveros-Palacios, N., Smith, A. M., Grundy, F. J., Soberon, M., Henkin, T. M., & Miranda-Ríos, J. (2008). Molecular basis of gene regulation by the THI-box riboswitch. *Molecular Microbiology*, *67*(4), 793–803. https://doi.org/10.1111/j.1365-2958.2007.06088.x

Otaka, H., Ishikawa, H., Morita, T., & Aiba, H. (2011a). PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(32), 13059–13064. https://doi.org/10.1073/pnas.1107050108

Otaka, H., Ishikawa, H., Morita, T., & Aiba, H. (2011b). PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(32), 13059–13064. https://doi.org/10.1073/pnas.1107050108

Otwinowski, Z., & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, *276*(January 1993), 307–326. https://doi.org/10.1016/S0076-6879(97)76066-X

Pandey, S., Gravel, C. M., Stockert, O. M., Wang, C. D., Hegner, C. L., LeBlanc, H., & Berry, K. E. (2020). Genetic identification of the functional surface for RNA binding by Escherichia coli ProQ. *Nucleic Acids Research*, *48*(8), 4507–4520. https://doi.org/10.1093/NAR/GKAA144

Papenfort, K., & Vanderpool, C. K. (2015). Target activation by regulatory RNAs in bacteria. In *FEMS Microbiology Reviews* (Vol. 39, Issue 3, pp. 362–378). Oxford University Press. https://doi.org/10.1093/femsre/fuv016

Patel, S., George, R., Autore, F., Fraternali, F., Ladbury, J. E., & Nikolova, P. V. (2008). Molecular interactions of ASPP1 and ASPP2 with the p53 protein family and the apoptotic promoters PUMA and Bax. *Nucleic Acids Research*, *36*(16), 5139–5151. https://doi.org/10.1093/nar/gkn490

Pelikan, M., Hura, G. L., & Hammel, M. (2009). Structure and flexibility within proteins as identified through small angle X-ray scattering. *General Physiology and Biophysics*, *28*, 174–189. https://doi.org/10.4149/gpb

Pinto, E. M., Musolino, N. R. C., Cescato, V. A. S., Soares, I. C., Wakamatsu, A., De Oliveira, E., Salgado, L. R., & Bronstein, M. D. (2010). IASPP: A novel protein involved in pituitary tumorigenesis? *Frontiers of Hormone Research*, *38*, 70–76. https://doi.org/10.1159/000318496

Plangger, R., Juen, M. A., Hoernes, T. P., Nußbaumer, F., Kremser, J., Strebitzer, E., Klingler, D., Erharter, K., Tollinger, M., Erlacher, M. D., & Kreutz, C. (2019). Branch site bulge conformations in domain 6 determine functional sugar puckers in group II intron splicing. *Nucleic Acids Research*, *47*(21), 11430–11440. https://doi.org/10.1093/nar/gkz965

Potts, A. H., Vakulskas, C. A., Pannuri, A., Yakhnin, H., Babitzke, P., & Romeo, T. (2017). Global role of the bacterial post-transcriptional regulator CsrA revealed by integrated transcriptomics. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-01613-1

Pourciau, C., Lai, Y. J., Gorelik, M., Babitzke, P., & Romeo, T. (2020). Diverse Mechanisms and Circuitry for Global Regulation by the RNA-Binding Protein CsrA. *Frontiers in Microbiology*, *11*(October). https://doi.org/10.3389/fmicb.2020.601352

Prozeller, D., Morsbach, S., & Landfester, K. (2019). Isothermal titration calorimetry as a complementary method for investigating nanoparticle-protein interactions. *Nanoscale*, *11*(41), 19265–19273. https://doi.org/10.1039/c9nr05790k

Putnam, C. D., Hammel, M., Hura, G. L., & Tainer, J. A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: Defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics*, *40*(3), 191–285. https://doi.org/10.1017/S0033583507004635

Quendera, A. P., Seixas, A. F., dos Santos, R. F., Santos, I., Silva, J. P. N., Arraiano, C. M., & Andrade, J. M. (2020). RNA-Binding Proteins Driving the Regulatory Activity of Small Non-coding RNAs in Bacteria. *Frontiers in Molecular Biosciences*, *7*(May), 1–9. https://doi.org/10.3389/fmolb.2020.00078

Ragusa, M. J., Dancheck, B., Critton, D. A., Nairn, A. C., Page, R., & Peti, W. (2010). Spinophilin directs protein phosphatase 1 specificity by blocking substrate binding sites. *Nature Structural and Molecular Biology*, *17*(4), 459–464. https://doi.org/10.1038/nsmb.1786

Rajkowitsch, L., Chen, D., Stampfl, S., Semrad, K., Waldsich, C., Mayer, O., Jantsch, M. F., Konrat, R., Bläsi, U., & Schroeder, R. (2007). RNA chaperones, RNA annealers and RNA helicases. *RNA Biology*, *4*(3), 118–130. https://doi.org/10.4161/rna.4.3.5445

Rambo, R. P., & Tainer, J. A. (2013). Super-resolution in solution x-ray scattering and its applications to structural systems biology. *Annual Review of Biophysics*, *42*(1), 415–441. https://doi.org/10.1146/annurev-biophys-083012-130301

Renda, A., Poly, S., Lai, Y., Pannuri, A., Yakhnin, H., Potts, A. H., Bevilacqua, P. C., Romeo, T., & Babitzke, P. (2020). CsrA-Mediated Translational Activation of ymdA Expression in Escherichia coli. *MBio*, *11*(5), e00849–20.

Reyes, F. E., Garst, A. D., & Batey, R. T. (2009). Strategies in RNA crystallography. In *Methods in enzymology* (1st ed., Vol. 469, Issue 09). Elsevier Inc. https://doi.org/10.1016/S0076-6879(09)69006-6

Rizvanovic, A., Kjellin, J., Söderbom, F., & Holmqvist, E. (2021). Saturation mutagenesis charts the functional landscape of Salmonella ProQ and reveals a gene regulatory function of its C-terminal domain. *Nucleic Acids Research*, *49*(17), 9992–10006. https://doi.org/10.1093/nar/gkab721

Robinson, R. A., Lu, X., Jones, E. Y., & Siebold, C. (2008). Biochemical and Structural Studies of ASPP Proteins Reveal Differential Binding to p53, p63, and p73. *Structure*, *16*(2), 259–268. https://doi.org/10.1016/j.str.2007.11.012

Rodionov, D. A., Vitreschak, A. G., Mironov, A. A., & Gelfand, M. S. (2002). Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms. *Journal of Biological Chemistry*, *277*(50), 48949–48959. https://doi.org/10.1074/jbc.M208965200

Rodolphe Barrangou, Christophe Fremaux, Hélène Deveau, Melissa Richards, Patrick Boyaval, Sylvain Moineau, Dennis A. Romero, & Philippe Horvath. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science*, *315*, 1709–1712. https://doi.org/10.1029/2004GL019460

Rogozin, I. B., Makarova, K. S., Natale, D. A., Spiridonov, A. N., Tatusov, R. L., Wolf, Y. I., Yin, J., & Koonin, E. v. (2002). Congruent evolution of different classes of non-coding DNA in prokaryotic genomes. *Nucleic Acids Research*, *30*(19), 4264–4271. www.ncbi.nlm.nih.gov:80/PMGifs/Genomes/org.html

Romeo, T., Gong, M., Mu Ya Liu, & Brun-Zinkernagel, A. M. (1993). Identification and molecular characterization of csrA, a pleiotropic gene from Escherichia coli that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. *Journal of Bacteriology*, *175*(15), 4744–4755. https://doi.org/10.1128/jb.175.15.4744-4755.1993

Rould, M. A., Perona, J. J., & Steitz, T. A. (1991). Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. *Nature*, *652*(6332), 213–218.

Samuels-Lev, Y., O'Connor, D. J., Bergamaschi, D., Trigiante, G., Hsieh, J. K., Zhong, S., Campargue, I., Naumovski, L., Crook, T., & Lu, X. (2001). ASPP proteins specifically stimulate the apoptotic function of p53. *Molecular Cell*, *8*(4), 781–794.

Sandercock, J. R., & Frost, L. S. (1998). Analysis of the major domains of the F fertility inhibition protein, FinO. *Molecular and General Genetics*, *259*(6), 622–629. https://doi.org/10.1007/s004380050856

Santiago-Frangos, A., Fröhlich, K. S., Jeliazkov, J. R., Małecka, E. M., Marino, G., Gray, J. J., Luisi, B. F., Woodson, S. A., & Hardwick, S. W. (2019). Caulobacter crescentus Hfq structure reveals a conserved mechanism of RNA annealing regulation. *Proceedings of the*

National Academy of Sciences of the United States of America, *166*(22), 10978–10987. https://doi.org/10.1073/pnas.1814428116

Santiago-Frangos, A., Jeliazkov, J. R., Gray, J. J., & Woodson, S. A. (2017). Acidic C-terminal domains autoregulate the RNA chaperone Hfq. *ELife*, *6*, 1–25. https://doi.org/10.7554/eLife.27049

Sauer, E., & Weichenrieder, O. (2011). Structural basis for RNA 3′-end recognition by Hfq. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(32), 13065–13070. https://doi.org/10.1073/pnas.1103420108

Sauter, C., Ng, J. D., Lorber, B., Keith, G., Brion, P., Hosseini, M. W., Lehn, J. M., & Giegé, R. (1999). Additives for the crystallization of proteins and nucleic acids. *Journal of Crystal Growth*, *196*(2–4), 365–376. https://doi.org/10.1016/S0022-0248(98)00852-5

Schlundt, A., Heinz, G. A., Janowski, R., Geerlof, A., Stehle, R., Heissmeyer, V., Niessing, D., & Sattler, M. (2014). Structural basis for RNA recognition in roquin-mediated post-transcriptional gene regulation. *Nature Structural and Molecular Biology*, *21*(8), 671–678. https://doi.org/10.1038/nsmb.2855

Schneidman-Duhovny, D., Hammel, M., Tainer, J. A., & Sali, A. (2016). FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Research*, *44*(W1), W424–W429. https://doi.org/10.1093/nar/gkw389

Schu, D. J., Zhang, A., Gottesman, S., & Storz, G. (2015). Alternative Hfq- sRNA interaction modes dictate alternative mRNA recognition . *The EMBO Journal*, *34*(20), 2557–2573. https://doi.org/10.15252/embj.201591569

Schubert, M., Lapouge, K., Duss, O., Oberstrass, F. C., Jelesarov, I., Haas, D., & Allain, F. H. T. (2007). Molecular basis of messenger RNA recognition by the specific bacterial repressing clamp RsmA/CsrA. *Nature Structural and Molecular Biology*, *14*(9), 807–813. https://doi.org/10.1038/nsmb1285

Schumacher, M. A., Pearson, R. F., Møller, T., Valentin-Hansen, P., & Brennan, R. G. (2002). Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: A bacterial Sm-like protein. *EMBO Journal*, *21*(13), 3546–3556. https://doi.org/10.1093/emboj/cdf322

Sgroi, D. C., Teng, S., Robinson, G., LeVangie, R., Hudson, J. R., & Elkahloun, A. G. (1999). In vivo gene expression profile analysis of human breast cancer progression. *Cancer Research*, *59*(22), 5656–5661.

Sherwood, A. V., & Henkin, T. M. (2016). Riboswitch-Mediated Gene Regulation: Novel RNA Architectures Dictate Gene Expression Responses. *Annual Review of Microbiology*, *70*, 361–374. https://doi.org/10.1146/annurev-micro-091014-104306

Shi, Y. (2009). Serine/Threonine Phosphatases: Mechanism through Structure. *Cell*, *139*(3), 468–484. https://doi.org/10.1016/j.cell.2009.10.006

Shopik, M. J., Li, L., Luu, H. A., Obeidat, M., Holmes, C. F. B., & Ballermann, B. J. (2013). Multi-directional function of the protein phosphatase 1 regulatory subunit TIMAP. *Biochemical and Biophysical Research Communications*, *435*(4), 567–573. https://doi.org/10.1016/j.bbrc.2013.05.012

Skene-Arnold, T. D., Luu, H. A., Uhrig, R. G., De Wever, V., Nimick, M., Maynes, J., Fong, A., James, M. N. G., Trinkle-Mulcahy, L., Moorhead, G. B., & Holmes, C. F. B. (2013). Molecular mechanisms underlying the interaction of protein phosphatase-1c with ASPP proteins. *Biochemical Journal*, *449*(3), 649–659. https://doi.org/10.1042/BJ20120506

Smirnov, A., Förstner, K. U., Holmqvist, E., Otto, A., Günster, R., Becher, D., Reinhardt, R., & Vogel, J. (2016). Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(41), 11591–11596. https://doi.org/10.1073/pnas.1609981113

Smirnov, A., Wang, C., Drewry, L. L., & Vogel, J. (2017). Molecular mechanism of mRNA repression in trans by a ProQ-dependent small RNA. *EMBO*, *36*, 1029–1045. https://doi.org/10.15252/embj.201696127

Smith, M. N., Crane, R. A., Keates, R. A. B., & Wood, J. M. (2004). Overexpression, purification, and characterization of ProQ, a posttranslational regulator for osmoregulatory transporter ProP of Escherichia coli. *Biochemistry*, *43*(41), 12979–12989. https://doi.org/10.1021/bi048561g

Someya, T., Baba, S., Fujimoto, M., Kawai, G., Kumasaka, T., & Nakamura, K. (2012). Crystal structure of Hfq from Bacillus subtilis in complex with SELEX-derived RNA aptamer: Insight into RNA-binding properties of bacterial Hfq. *Nucleic Acids Research*, *40*(4), 1856–1867. https://doi.org/10.1093/nar/gkr892

Soper, T. J., Doxzen, K., & Woodson, S. A. (2011). Major role for mRNA binding and restructuring in sRNA recruitment by Hfq. *Rna*, *17*(8), 1544–1550. https://doi.org/10.1261/rna.2767211

Soubeyrand, S., Schild-Poulter, C., & Haché, R. J. G. (2004). Structured DNA promotes phosphorylation of p53 by DNA-dependent protein kinase at serine 9 and threonine 18. *European Journal of Biochemistry*, *271*(18), 3776–3784. https://doi.org/10.1111/j.1432-1033.2004.04319.x

Sowa, S. W., Gelderman, G., Leistra, A. N., Buvanendiran, A., Lipp, S., Pitaktong, A., Vakulskas, C. A., Romeo, T., Baldea, M., & Contreras, L. M. (2017). Integrative FourD omics approach profiles the target network of the carbon storage regulatory system. *Nucleic Acids Research*, *45*(4), 1673–1686. https://doi.org/10.1093/nar/gkx048

Stanek, K. A., Patterson-West, J., Randolph, P. S., & Mura, C. (2017). Crystal structure and RNA-binding properties of an Hfq homolog from the deep-branching Aquificae: Conservation of the lateral RNA-binding mode. *Acta Crystallographica Section D: Structural Biology*, *73*(4), 294–315. https://doi.org/10.1107/S2059798317000031

Stein, E. M., Kwiatkowska, J., Basczok, M. M., Gravel, C. M., Berry, K. E., & Olejniczak, M. (2020). Determinants of RNA recognition by the FinO domain of the Escherichia coli ProQ protein. *Nucleic Acids Research*, *48*(13), 7502–7519. https://doi.org/10.1093/nar/gkaa497

Stougaard, P., Molin, S., & Nordstrom, K. (1981). RNAs involved in copy-number control and incompatibility of plasmid R1. *Proceedings of the National Academy of Sciences of the United States of America*, *78*(10 I), 6008–6012. https://doi.org/10.1073/pnas.78.10.6008

Studer, G., Rempfer, C., Waterhouse, A. M., Gumienny, R., Haas, J., & Schwede, T. (2020). QMEANDisCo—distance constraints applied on model quality estimation. *Bioinformatics*, *36*(6), 1765–1771. https://doi.org/10.1093/bioinformatics/btz828

Sudarsan, N., Barrick, J. E., & Breaker, R. R. (2003). Metabolite-binding RNA domains are present in the genes of eukaryotes. *Rna*, *9*(6), 644–647. https://doi.org/10.1261/rna.5090103

Tan, D., Zhou, M., Kiledjian, M., & Tong, L. (2014). The ROQ domain of Roquin recognizes mRNA constitutive-decay element and double-stranded RNA. *Nature Structural and Molecular Biology*, *21*(8), 679–685. https://doi.org/10.1038/nsmb.2857

Tatum, E. L., & Lederberg, J. (1946). Gene recombination in Escherichia coli. *Nature*, *158*, 558.

Teplova, M., Song, J., Gaw, H. Y., Teplov, A., & Patel, D. J. (2010). Structural Insights into RNA Recognition by the Alternate-Splicing Regulator CUG-Binding Protein 1. *Structure*, *18*(10), 1364–1377. https://doi.org/10.1016/j.str.2010.06.018

Teplova, M., Yuan, Y. R., Phan, A. T., Malinina, L., Ilin, S., Teplov, A., & Patel, D. J. (2006). Structural basis for recognition and sequestration of UUUOH 3′ temini of nascent RNA polymerase III transcripts by La, a rheumatic disease autoantigen. *Molecular Cell*, *21*(1), 75–85. https://doi.org/10.1016/j.molcel.2005.10.027

Terrak, M., Kerff, F., Langsetmo, K., & Tao, T. (2004). Structural basis of protein phosphatase 1 regulation. *Nature*, *429*, 780–784.

Timmis, K. N., Andres, I., & Achtman, M. (1978). Fertility repression of F-like conjugative plasmids: Physical mapping of the R6-5 finO and finP cistrons and identification of the finO protein. *Proceedings of the National Academy of Sciences of the United States of America*, *75*(12), 5836–5840. https://doi.org/10.1073/pnas.75.12.5836

Timmis, K. N., Cabello, F., & Cohen, S. N. (1978). Cloning and characterization of EcoRI and HindIII restriction endonuclease-generated fragments of antibiotic resistance plasmids R6-5 and R6. *MGG Molecular & General Genetics*, *162*(2), 121–137. https://doi.org/10.1007/BF00267869

Tomizawa, J. ichi, & Som, T. (1984). Control of cole 1 plasmid replication: Enhancement of binding of RNA I to the primer transcript by the rom protein. *Cell*, *38*(3), 871–878. https://doi.org/10.1016/0092-8674(84)90282-4

Tomizawa, J., Itoh, T., Selzer, G., & Som, T. (1981). Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA. *Proceedings of the National Academy of Sciences of the United States of America*, *78*(3), 1421–1425. https://doi.org/10.1073/pnas.78.3.1421

Tree, J. J., Granneman, S., McAteer, S. P., Tollervey, D., & Gally, D. L. (2014). Identification of Bacteriophage-Encoded Anti-sRNAs in Pathogenic Escherichia coli. *Molecular Cell*, *55*(2), 199–213. https://doi.org/10.1016/j.molcel.2014.05.006

Trigiante, G., & Lu, X. (2006). ASPPs and cancer. *Nature Reviews Cancer*, *6*(3), 217–226. https://doi.org/10.1038/nrc1818

Uchikawa, E., Natchiar, K. S., Han, X., Proux, F., Roblin, P., Zhang, E., Durand, A., Klaholz, B. P., & Dock-Bregeon, A. C. (2015). Structural insight into the mechanism of stabilization of the 7SK small nuclear RNA by LARP7. *Nucleic Acids Research*, *43*(6), 3373–3388. https://doi.org/10.1093/nar/gkv173

Valverde, C., Lindell, M., Wagner, E. G. H., & Haas, D. (2004). A repeated GGA motif is critical for the activity and stability of the riboregulator RsmY of Pseudomonas fluorescens. *Journal of Biological Chemistry*, *279*(24), 25066–25074. https://doi.org/10.1074/jbc.M401870200

van Biesen, T., Soderbom, F., Wagner, E. G. H., & Frost, L. S. (1993). Structural and functional analyses of the FinP antisense RNA regulatory system of the F conjugative piasmid. *Molecular Microbiology*, *10*(1), 35–43. https://doi.org/10.1111/j.1365-2958.1993.tb00901.x

Vigneron, A. M., Ludwig, R. L., & Vousden, K. H. (2010). Cytoplasmic ASPP1 inhibits apoptosis through the control of YAP. *Genes and Development*, *24*(21), 2430–2439. https://doi.org/10.1101/gad.1954310

Vogel, J., & Luisi, B. F. (2011). Hfq and its constellation of RNA. *Nature Reviews Microbiology*, *9*(8), 578–589. https://doi.org/10.1038/nrmicro2615

Wang, L., Xing, H., Tian, Z., Peng, L., Li, Y., Tang, K., Rao, Q., Wang, M., & Wang, J. (2012). IASPPsv antagonizes apoptosis induced by chemotherapeutic agents in MCF-7 cells and mouse thymocytes. *Biochemical and Biophysical Research Communications*, *424*(3), 414–420. https://doi.org/10.1016/j.bbrc.2012.06.124

Wang, W., Wang, L., Wu, J., Gong, Q., & Shi, Y. (2013). Hfq-bridged ternary complex is important for translation activation of rpoS by DsrA. *Nucleic Acids Research*, *41*(11), 5938–5948. https://doi.org/10.1093/nar/gkt276

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–W303. https://doi.org/10.1093/nar/gky427

Waters, L. S., & Storz, G. (2009). Regulatory RNAs in Bacteria. *Cell*, *136*(4), 615–628. https://doi.org/10.1016/j.cell.2009.01.043

Watkins, D., & Arya, D. P. (2019). Regulatory roles of small RNAs in prokaryotes: parallels and contrast with eukaryotic miRNA. *Non-Coding RNA Investigation*, *3*, 28–28. https://doi.org/10.21037/ncri.2019.10.02

Weilbacher, T., Suzuki, K., Dubey, A. K., Wang, X., Gudapaty, S., Morozov, I., Baker, C. S., Georgellis, D., Babitzke, P., & Romeo, T. (2003). A novel sRNA component of the carbon storage regulatory system of Escherichia coli. *Molecular Microbiology*, *48*(3), 657–670. https://doi.org/10.1046/j.1365-2958.2003.03459.x

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., & Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods in Molecular Biology (Clifton, N.J.)*, *112*, 531–552. https://doi.org/10.1385/1-59259-584-7:531

Wimberly, B. T., Brodersen, D. E., Clemons Jr, W. M., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartschk, T., & Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. In *NATURE* (Vol. 407). www.nature.com

Winkler, W., Nahvi, A., & Breaker, R. R. (2002). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, *419*(6910), 952–956.

Yakhnin, A. v., Baker, C. S., Vakulskas, C. A., Yakhnin, H., Berezin, I., Romeo, T., & Babitzke, P. (2013). CsrA activates flhDC expression by protecting flhDC mRNA from RNase E-mediated cleavage. *Molecular Microbiology*, *87*(4), 851–866. https://doi.org/10.1111/mmi.12136

Yang, J. P., Hori, M., Sanda, T., & Okamoto, T. (1999). Identification of a novel inhibitor of nuclear factor-κB, RelA- associated inhibitor. *Journal of Biological Chemistry*, *274*(22), 15662–15670. https://doi.org/10.1074/jbc.274.22.15662

Yang, J. P., Hori, M., Takahashi, N., Kawabe, T., Kato, H., & Okamoto, T. (1999). NF-κB subunit p65 binds to 53BP2 and inhibits cell death induced by 53BP2. *Oncogene*, *18*(37), 5177–5186. https://doi.org/10.1038/sj.onc.1202904

Zhang, B., Xiao, H. J., Chen, J., Tao, X., & Cai, L. H. (2011). Inhibitory member of the apoptosis-stimulating protein of p53 (ASPP) family promotes growth and tumorigenesis in human p53-deficient prostate cancer cells. *Prostate Cancer and Prostatic Diseases*, *14*(3), 219–224. https://doi.org/10.1038/pcan.2011.25

Zhang, X., Wang, M., Zhou, C., Chen, S., & Wang, J. (2005). The expression of iASPP in acute leukemias. *Leukemia Research*, *29*(2), 179–183. https://doi.org/10.1016/j.leukres.2004.07.001

Zhao, J., Wu, G., Bu, F., Lu, B., Liang, A., Cao, L., Tong, X., Lu, X., Wu, M., & Guo, Y. (2010). Epigenetic silence of ankyrin-repeat-containing, SH3-domain-containing, and proline-rich-region-containing protein 1 (ASPP1) and ASPP2 genes promotes tumor growth in hepatitis B virus-positive hepatocellular carcinoma. *Hepatology*, *51*(1), 142–153. https://doi.org/10.1002/hep.23247

Zheng, A., Panja, S., & Woodson, S. A. (2016). Arginine Patch Predicts the RNA Annealing Activity of Hfq from Gram-Negative and Gram-Positive Bacteria. *Journal of Molecular Biology*, *428*(11), 2259–2264. https://doi.org/10.1016/j.jmb.2016.03.027

Zhou, Y., Millott, R., Kim, H. J., Peng, S., Edwards, R. A., Skene-Arnold, T., Hammel, M., Lees-Miller, S. P., Tainer, J. A., Holmes, C. F. B., & Glover, J. N. M. (2019). Flexible Tethering of ASPP Proteins Facilitates PP-1c Catalysis. *Structure*, *27*(10), 1485-1496.e4. https://doi.org/10.1016/j.str.2019.07.012

Zhuo, S., Clemens, J. C., Hakes, D. J., Barford, D., & Dixon, J. E. (1993). Expression, purification, crystallization, and biochemical characterization of a recombinant protein phosphatase. *Journal of Biological Chemistry*, *268*(24), 17754–17761. https://doi.org/10.1016/s0021-9258(17)46769-x

Appendix A.

The investigation of RocC chaperone activity and the discovery of RocC:RocR:*comEA* mRNA trimer complex

A.1. Overview

RocC and RocR are key components to repress competence in *L. pneumophila*. An *in vivo* assay revealed that lack of either one of the components failed to suppress natural transformation. The ProQ/FinO family is reported to facilitate chaperone activity. RocC was assumed to have chaperone activity due to its possession of a ProQ/FinO domain however any chaperone activity associated with RocC had not been tested. RNA chaperone activity involves the ability to destabilize internal base-pairing and promote annealing of two RNAs. In this appendix, a set of experiments were performed to understand the ability of RocC to promote annealing of RocR and *comEA* mRNA 5' UTR, one of RocR's known targets. The annealing assay revealed RocR and *comEA* mRNA forms a complex without RocC assistance. However, we did observe that RocC could participate in a 2RocC:2RocR complex, suggesting the possibility that RocC may exhibit some abililty to facilitate RocR-RocR association. In addition, EMSA revealed a potential RocC:RocR:*comEA* mRNA complex.

A.2. Introduction

The RNA chaperone activity of the ProQ/FinO family was demonstrated using FinP and its complementary RNA *traJ* mRNA with *E. coli* FinO, ProQ, and *N. meningitidis* NMB1681 (Chaulk et al., 2010, 2011; Ghetu et al., 2000). RNA chaperones are generally thought to remodel RNAs by disrupting internal base-pairs and promoting new base pairings. In the FinOP system, the ability to melt internal hydrogen bonds and enhance duplex formation were demonstrated with strand exchange and RNA annealing (or duplexing) assays (see Chapter 1 and Figure 1-7B, C) (Arthur et al., 2003). RocC is a member of the ProQ/FinO family in *L. pneumophila* and has been shown to be involved in a system that represses the expression of genes required for the uptake of environmental DNA, known as natural competence (Attaiech et al., 2016). RocR, the primary RNA target of RocC, is composed of 3 stem-loop structures (SLs) of which the final, SL3 is part of the transcription terminator. EMSA revealed strong binding to SL3 of RocR through the RocC ProQ/FinO domain, and SL1 and SL2 were predicted to base-pair with 4 target mRNAs to regulate competence development. In addition, RocC was also shown to enhance the half-life of RocR (Attaiech et al., 2016).

The 4 target mRNAs share a conserved seed region called the RocR box near the RBS (Attaiech et al., 2016). These 6 nucleotides are complementary to the exposed loop in SL1. Mutations in the RocR box reduced the ability of RocR to repress gene expression, suggesting RocR:*comEA* mRNA hybridization is critical for gene regulation. *In vivo* analysis demonstrated that RocR:*comEA* mRNA base-pairing and RocC interaction with RocR increases RocR stability and is essential for gene repression (Attaiech et al., 2016). However, it remained unclear whether RocC could remodel RNA structure or could facilitate pairing between RocR and *comEA* mRNA.

Here, we attempted to understand the ability of RocC to promote annealing of RNAs using EMSA. RNA annealing assays showed that RocR and *comEA* mRNA 5' UTR spontaneously forms an unstable complex in the absence of RocC. We found that RocC binds through SL3 and potentially promotes self-dimerization of RocR. RocC-facilitated RocR dimerization required the RocC C-terminal domain. We also found a potential trimeric complex composed of RocC, RocR and *comEA* mRNA.

A.3. Results

A.3.1. RocR and *comEA* mRNA can form a complex that is sufficiently stable to be observed in EMSA

We had previously used EMSA to measure the ability of the complementary RNA pair FinP and *traJ* to interact, and to demonstrate that this pairing is facilitated by FinO (Arthur et al., 2003). In this case, FinP and *traJ* are perfectly complementary, however, the presence of hairpin structures within both RNAs served as kinetic blocks against their hybridization. To test whether RocC also has RNA chaperone activity that facilitates RocR interaction with *comEA* mRNA, we wanted to first test if the limited pairing that is expected to form between these two RNAs is sufficient to make a stable complex in EMSA. To test this, we labeled RocR with ^{32}P, and tested its ability to interact with a molar excess of unlabeled RNA corresponding to the 5' UTR of *comEA* mRNA by non-denaturing gel electrophoresis (EMSA). Our results suggest that the presence of the *comEA* mRNA does not impact the mobility of RocR if the sample is run on the gel immediately after mixing. However, if the mixture is allowed to equilibrate, a lower mobility species is observed and this species becomes more prominent as the mixing time is increased (Figure A-1). This species was also observed if the mixture was denatured at 90°C followed by renaturation by either slow cooling or snap cooling. We suggest that this species corresponds to a RocR:*comEA* complex, however the fact that not all the RocR is shifted into this complex may suggest that this complex is not kinetically stable enough for all the sample to remain paired through the EMSA assay. Nevertheless, this experiment demonstrates that enough complex can be detected so that EMSA can be used to monitor interactions between these RNAs.

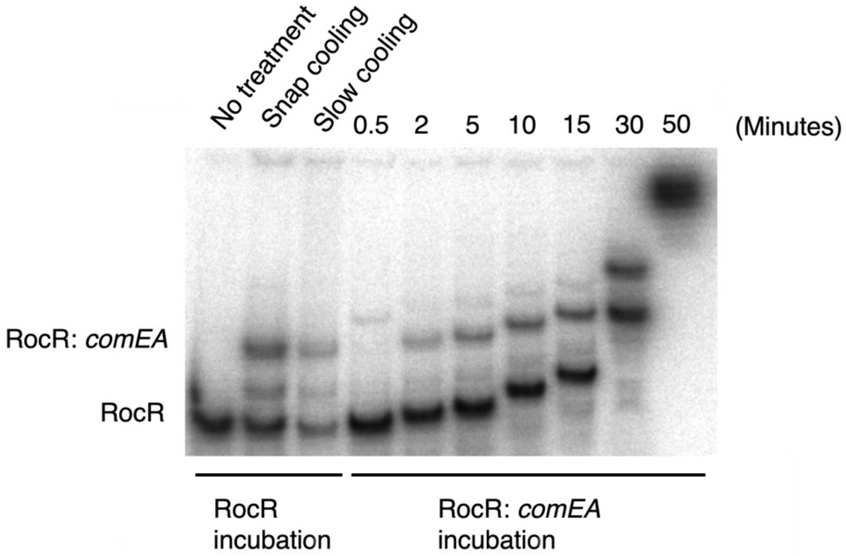


Figure A-1. RNA annealing assay and SDS concentration test.

(A) Testing annealing of 5’ radiolabeled RocR and *comEA* mRNA 5’ UTR over time by 10% non-denaturing gel electrophoresis.

A.3.2. Full-length RocC can form two distinct complexes with RocR

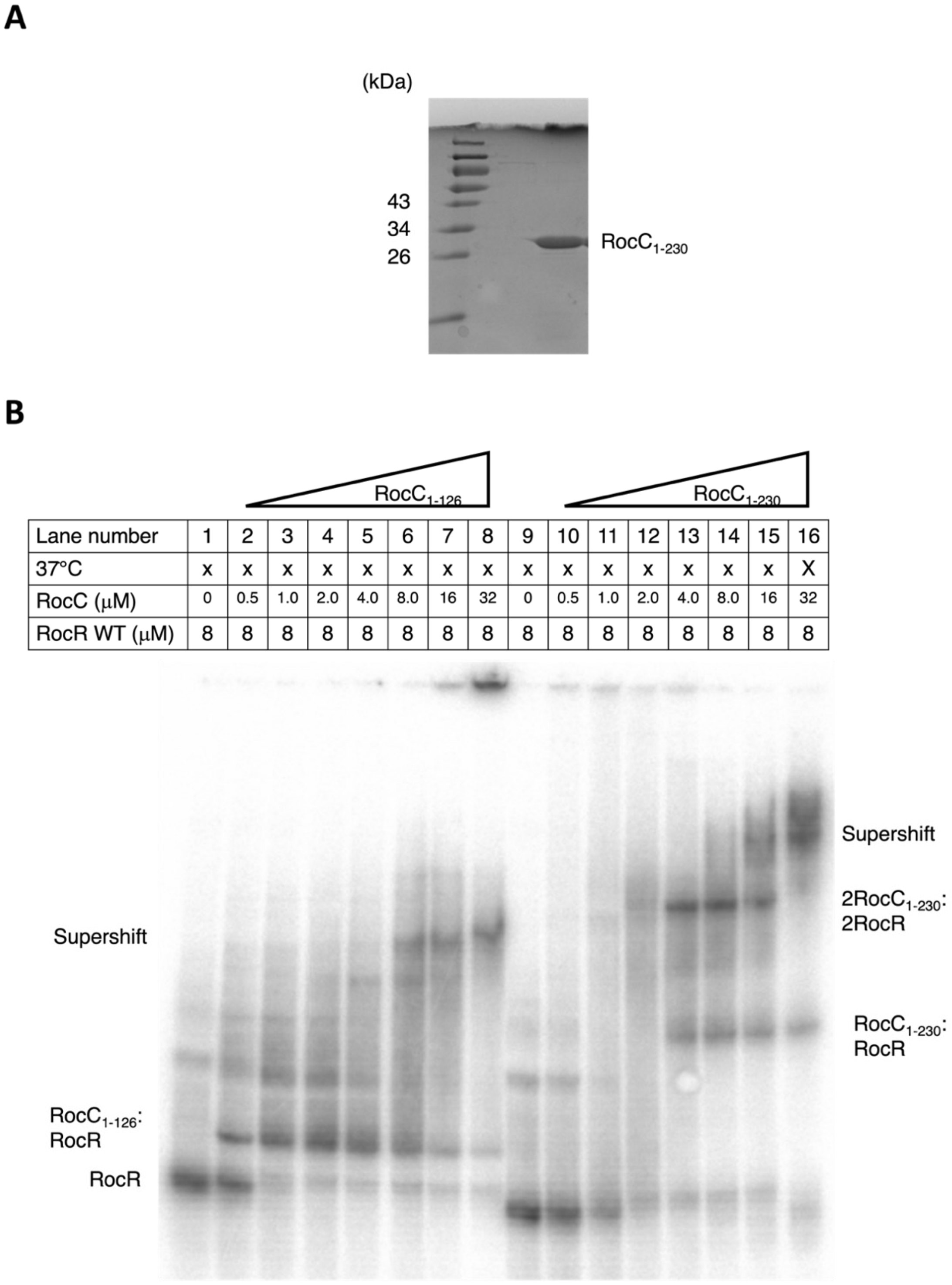
We next used EMSA to further probe interactions of RocC with RocR. Our preliminary work previously demonstrated that RocC can bind RocR with ~µM binding affinity, and that most of this interaction is due to recognition of the 3’ SL3 of RocR with the RocC ProQ/FinO domain (Attaiech et al., 2016). We revisited these experiments with more highly purified RocC and found that actually RocC forms two distinct complexes with RocR. SDS-PAGE analysis of our RocC indicates that the two species are not simply due to degradation of the RocC protein (Figure A-2A). We therefore conclude that RocC can bind RocR to form two stoichiometrically-distinct complexes (Figure A-2B). We also compared this interaction to the interaction of the isolated RocC FinO domain (RocC$_{1\text{-}126}$). While these experiments revealed significant non-specific binding, the results suggest this protein binds RocR to yield only a single defined complex, and not the two complexes formed by the full-length protein.

A.3.3. RocC can form a complex with RocR and *comEA* RNA

One possible way in which RocC might facilitate RocR repression of *comEA* could be through direct stabilization of RocR:*comEA* RNA interactions. To ascertain whether RocC might interact with the RocR:*comEA* RNA complex, we titrated RocC into a mixture of RocR and *comEA* RNA and separated the resulting species by EMSA (Figure A-3). The results appear very similar to the control titration of RocC against RocR alone. However, we do observe some differences. The band corresponding to the RocR:*comEA* RNA complex disappears as RocC is added and a new faint band of lower mobility is observed, which we suggest corresponds to a RocC:RocR:*comEA* RNA complex. To further validate this notion, we titrated increasing concentrations of *comEA* RNA into a RocC:RocR mixture (Figure A-4). These results show that with increasing concentrations of *comEA* RNA, the intensity of the tentative RocC:RocR:*comEA* RNA band increases. In addition, the lower mobility of the two RocC:RocR complexes reduces in intensity, suggesting that this complex but not the higher mobility RocC:RocR complex seems to preferentially bind *comEA* RNA.

A.4. Discussion

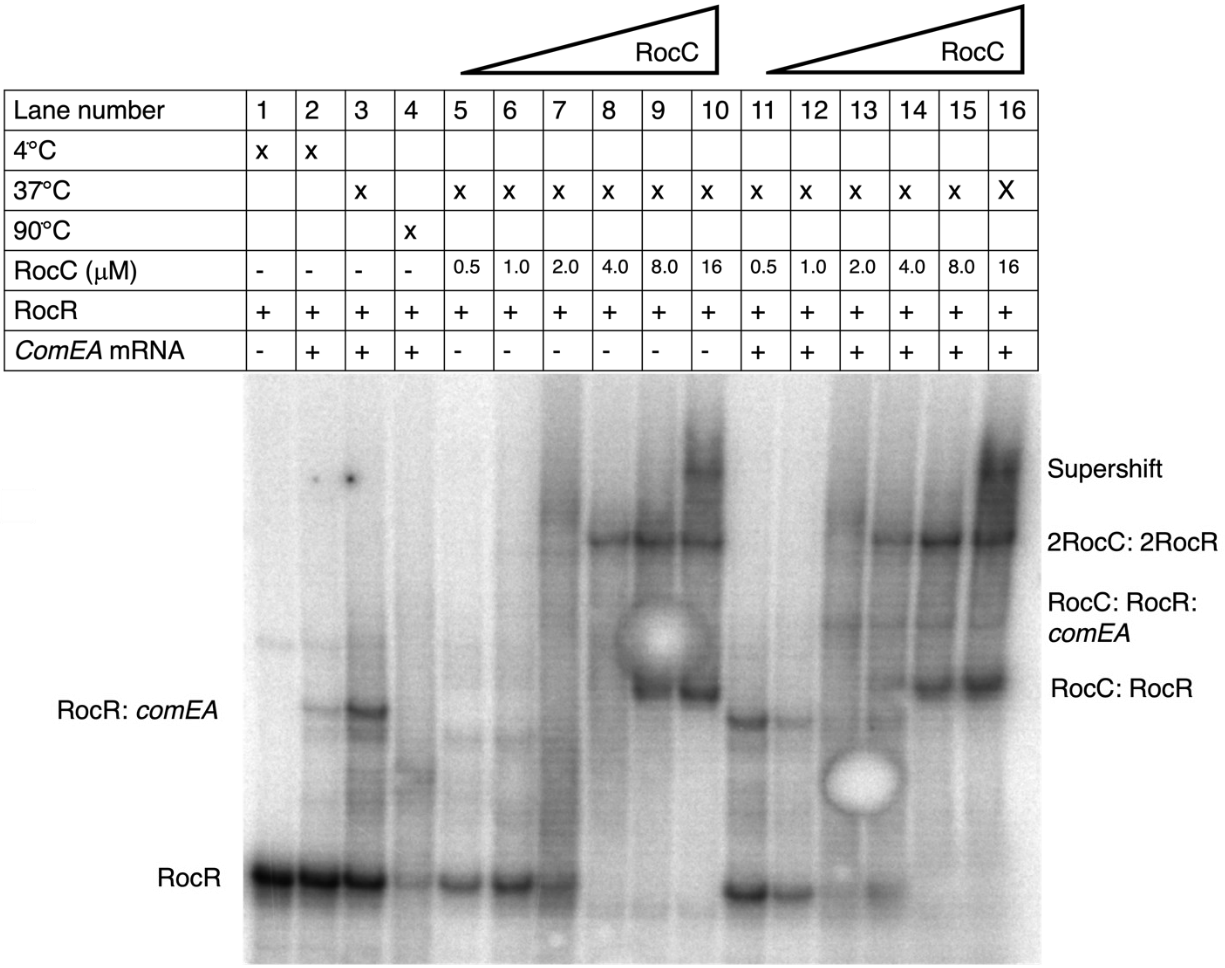
While most of the study of FinO domain proteins has focused on their ability to specifically bind 3’ RNA structures, it is also clear that this function is not sufficient for biological activity. This has been clearly demonstrated for FinO and also for RocC. For RocC, truncations of the C-terminal domain ablate the ability of RocC to repress competence in vivo, even though the truncated protein is stable in cells and can stabilize RocR (Attaiech et al., 2016). Thus, there must be some additional critical activity that facilitates function. For FinO, we believe this function is associated with the protein-assisted destabilization of internal RNA secondary structures that can



Lane number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
37°C	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	X
RocC (μM)	0	0.5	1.0	2.0	4.0	8.0	16	32	0	0.5	1.0	2.0	4.0	8.0	16	32
RocR WT (μM)	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8

Figure A-2. EMSA to test the impact of the RocC C-terminus on formation of the higher order RocC:RocR complex.

(A) Purified RocC$_{1\text{-}230}$ on 15% SDS-PAGE gel. (B) RocR was 5’ radiolabeled and visualized on a 10% native gel. Temperature indicates the sample incubation temperature. The concentration of RocC in each lane is indicated.

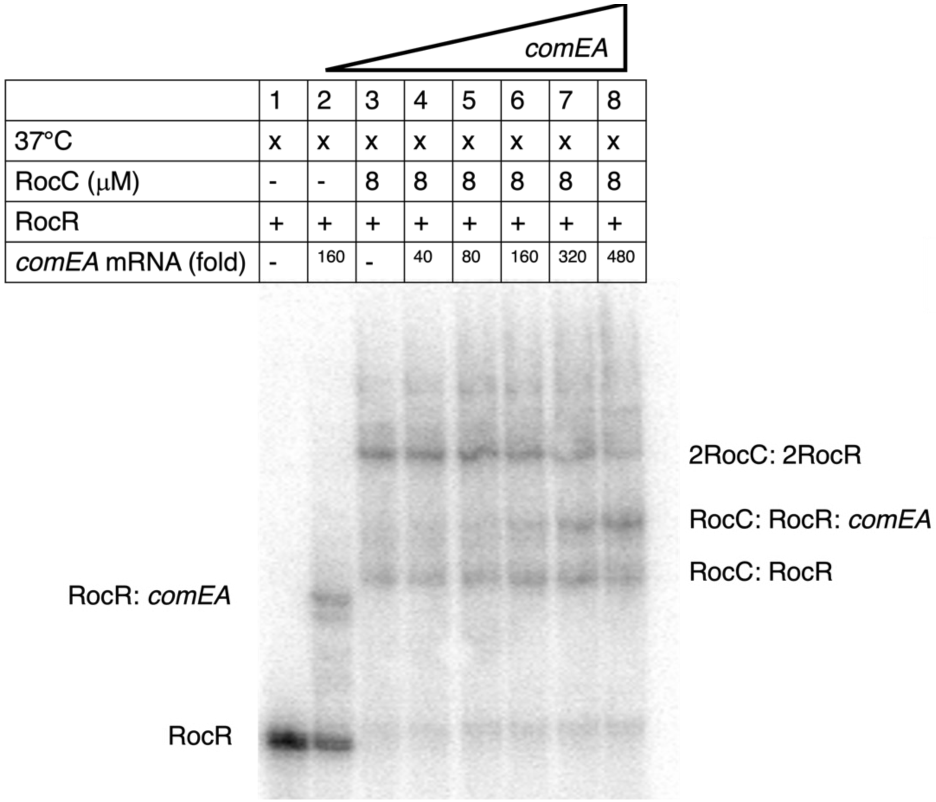


Lane number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
4°C	x	x														
37°C			x		x	x	x	x	x	x	x	x	x	x	x	X
90°C				x												
RocC (μM)	-	-	-	-	0.5	1.0	2.0	4.0	8.0	16	0.5	1.0	2.0	4.0	8.0	16
RocR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ComEA mRNA	-	+	+	+	-	-	-	-	-	-	+	+	+	+	+	+

Figure A-3. EMSA to test the role of RocC in RocR:*comEA* mRNA annealing.

RocR WT was 5’ radiolabeled and visualized on the 10% native gel. Temperature indicates the sample pre-incubation temperature.

then facilitate RNA:RNA duplexing. In the case of the RocC:RocR:*comEA* system, a completely complementary duplex is not possible, and interactions between these RNAs likely only involve limited base pairing. Indeed, the results shown here do indicate that RocR and *comEA* RNA can interact, however these interactions are likely less stable in gel electrophoresis than the FinP:*traJ* RNA duplex. Our results also suggest that interactions between full-length RocC and RocR are more complex than the interactions of the isolated RocC FinO domain with the RocR terminator. RocC appears to interact with RocR in a way that involves interaction of multiple RocC molecules with RocR, or else that RocC interaction triggers RocR oligomerization. This species appears to be able to interact with *comEA* RNA to form a new complex. Interestingly, this complex has a higher mobility than the RocC:RocR complex, suggesting that it may be smaller than the RocC:RocR complex.



	1	2	3	4	5	6	7	8
37°C	x	x	x	x	x	x	x	x
RocC (μM)	-	-	8	8	8	8	8	8
RocR	+	+	+	+	+	+	+	+
comEA mRNA (fold)	-	160	-	40	80	160	320	480

Figure A-4. EMSA to test the formation of the RocC:RocR:comea mRNA trimer.

RocC was incubated with 5’ radiolabeled RocR at 37°C. The number of *comEA* mRNA is the fold molar excess over RocR.

We suggest that the RocC:RocR complex might be an heterotetramer (Figure A-5A, A-6). The heterotetramer could be in either of two forms (Figure A-6). One possibility could be where the two RocR RNAs interact through limited imperfect base pairing between the two RNAs while the RocC proteins bind to the 3’ transcription terminator hairpin-tail structures (Figure A-6A). A second possibility would involve the unwinding of SL3 and the pairing of the SL3 regions of the two RocR molecules to stabilize RocR:RocR complex (Figure A-6B). We previously showed that

FinO, ProQ and NMB1681 can all facilitate duplexing between stem-tail RNA terminator structures and it seems likely that this could also occur with RocC. Our structural work shows that RocC ProQ/FinO domain interacts almost exclusively with the 3' strand of the terminator hairpin, presenting the possibility that the hairpin could unwind and pair with a complementary strand with RocC remaining bound. This would leave the remaining 5' region of RocR free to interact with *comEA* RNA. The fact that the RocC:RocR:*comEA* RNA complex has a higher mobility than the RocC:RocR complex suggests that formation of the former likely involves dissociation of the RocC:RocR hetero-tetramer.

Further work is required to develop this model. First, experiments with a higher concentration of *comEA* RNA should be performed to better visualize complexes involving this RNA. Second, we will do further work to determine the role of the C-terminal domain of RocC in this process. Is the C-terminal domain truly required to allow formation of a RocC:RocR:*comEA* complex? Can the C-terminal domain facilitate formation of a high oligomeric complex with just the RocR hairpin-tail terminator structure? This latter experiment could directly distinguish between the two models of hetero-tetramer presented in Figure A-5. If the C-terminal domain is truly critical for this activity, it would be interesting to mutagenize this region to gain more details into what residues are truly important for this activity. It would then be interesting to test the effects of these mutations *in vivo* to see if there is a good correlation between the ability to complex with *comEA*, and repression of *comEA* in living cells.

Finally, structural studies involving larger RocC constructs might yield mechanistic insights into the basis of RocC chaperone activity. Previous work on the ProQ system suggested that the C-terminal domain of that protein does contact its RNA targets (Gonzalez et al., 2017). In the FinO system, there is good evidence that it is the N-terminal unstructured region that is particularly critical for chaperone function. While that region is dispensable for high affinity RNA recognition, FRET and RNA-protein crosslinking demonstrated that that domain does in fact contact RNA.

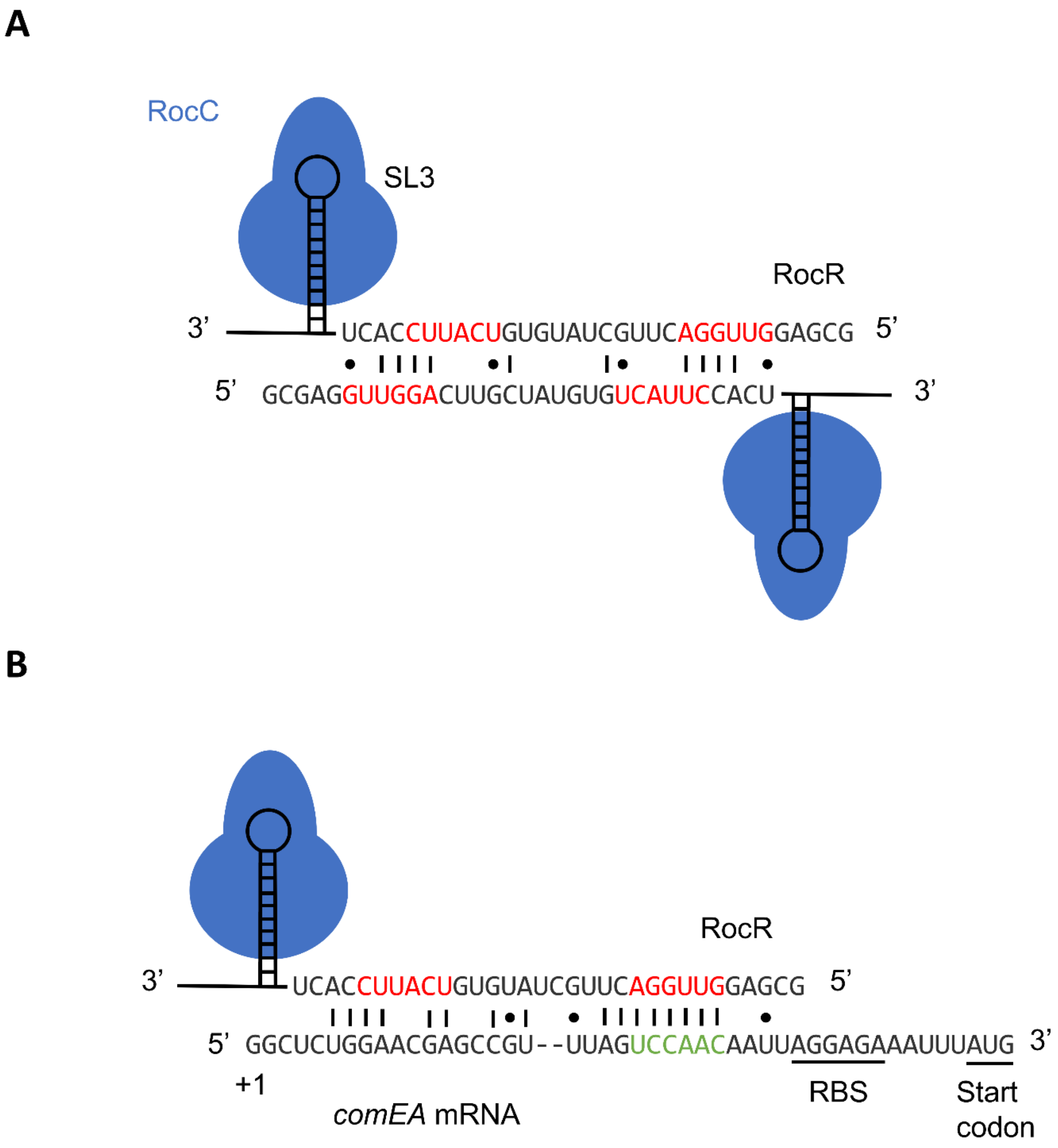


Figure A-5. Predicted RocC:RocR 2:2 complex and RocC:RocR:*comEA* mRNA trimer complex.

(A) Potential RocC:RocR 2:2 complex. Red is the loop of SL1 and SL2. Lines are Watson-Crick base pairs, and dots are wobble base pairs. (B) Potential RocC:RocR:*comEA* mRNA trimer complex. Green is RocR box.

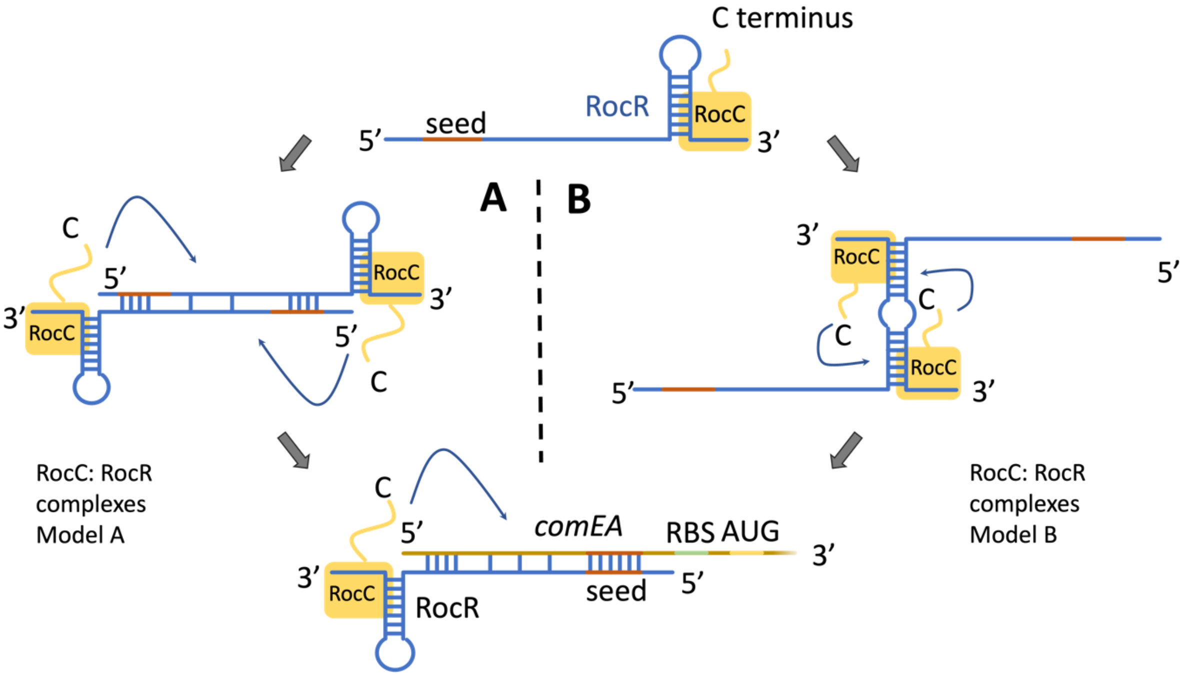


Figure A-6. Two possible RocC:RocR heterotetramer models.

(A) Individual RocC binds through SL3 of RocR, and two RocR forms duplex through SL1 and SL2. (B) Duplex forms through SL3 hybridization. RocC binds to RocR through SL3-like hybridized-duplex stem.

A.5. Materials and Methods

EMSA

EMSA was performed with minor modifications from Chapter 2.

RNA annealing assay

The 5’ radiolabeled RocR and non-labeled *comEA* mRNA complex for titrating with RocC gradient was prepared by incubating at least a 10-fold molar excess of *comEA* mRNA was incubated with RocR. Annealed RNAs by heat were incubated at 90°C at the heating block for two minutes. After incubation, the sample, which cool down in the ice, is called snap cooling. In the slow cooling sample, heating was turned off after two minutes incubation and was left until the heating block came down to room temperature.

Labeled RocR (450 nM in 40 µL) were incubated 50 minutes at 37°C with a 20-fold higher molar excess of *comEA* mRNA (8.8 µM in 40µL) in 40 µL EMSA reaction buffer (25 mM HEPES pH 7.3, 150 mM NaCl, 4 mM $MgCl_2$, 10% glycerol, 1 mM DTT, and 12 U RNaseOUT (ThermoFisher)). Each time point from 0.5 to 50 minutes, 5µL of the sample was taken from the mixture and mixed with 1 µL of 5x native gel loading dye (10 mM Tris pH 8.0, 50% glycerol, 0.001% bromophenol blue, 0.001% Xylene cyanol FF). 4 µL out of 6 µL were loaded into continuously running 10% native gel. The running condition was the same as EMSA in chapter 2.

A.6. References

Arthur, D. C., Ghetu, A. F., Gubbins, M. J., Edwards, R. A., Frost, L. S., & Glover, J. N. M. (2003). FinO is an RNA chaperone that facilitates sense-antisense RNA interactions. *EMBO Journal*, *22*(23), 6346–6355. https://doi.org/10.1093/emboj/cdg607

Attaiech, L., Boughammoura, A., Brochier-Armanet, C., Allatif, O., Peillard-Fiorente, F., Edwards, R. A., Omar, A. R., MacMillan, A. M., Glover, M., & Charpentier, X. (2016). Silencing of natural transformation by an RNA chaperone and a multitarget small RNA. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 8813–8818. https://doi.org/10.1073/pnas.1601626113

Chaulk, S., Lu, J., Tan, K., Arthur, D. C., Edwards, R. a, Frost, L. S., Joachimiak, A., & Glover, J. N. M. (2010). N. meningitidis 1681 is a member of the FinO family of RNA chaperones. *RNA Biology*, *7*(6), 812–819. https://doi.org/10.4161/rna.7.6.13688

Chaulk, S., Smith-Frieday, M. N., Arthur, D. C., Culham, D. E., Edwards, R. A., Soo, P., Frost, L. S., Keates, R. A. B., Glover, J. N. M., & Wood, J. M. (2011). ProQ is an RNA chaperone that controls ProP levels in escherichia coli. *Biochemistry*, *50*(15), 3095–3106. https://doi.org/10.1021/bi101683a

Ghetu, A. F., Gubbins, M. J., Frost, L. S., & Glover, J. N. M. (2000). Crystal structure of the bacterial conjugation repressor FinO. *Nature Structural Biology*, *7*(7), 565–569. https://doi.org/10.1038/76790

Gonzalez, G. M., Hardwick, S. W., Maslen, S. L., Mark Skehel, J., Holmqvist, E., Vogel, J., Bateman, A., Luisi, B. F., & William Broadhurst, R. (2017). Structure of the Escherichia coli ProQ RNA-binding protein. *RNA*, *23*(5), 696–711. https://doi.org/10.1261/rna.060343.116

Appendix B.

Flexible tethering of ASPP proteins facilitates PP-1c catalysis

This chapter was published:

Yeyun Zhou†, Robyn Millott†, Hyeong Jin Kim†, ShiyunPeng, Ross A. Edwards, Tamara Skene-Arnold, Michal Hammel, Susan P.Lees-Miller, John A. Tainer, Charles F. B. Holmes, J. N. Mark Glover. Flexible Tethering of ASPP Proteins Facilitates PP-1c Catalysis. *Structure*.

† These authors contributed equally.

Hyeong Jin Kim performed EMSA, SEC-SAXS and analyzed performed EMSA, SEC-SAXS data.

B.1. Overview

ASPP (apoptosis-stimulating proteins of p53) proteins bind PP-1c (protein phosphatase 1) and regulate p53 impacting cancer cell growth and apoptosis. Here we determine the crystal structure of the oncogenic ASPP protein, iASPP, bound to PP-1c. The structure reveals a 1:1 complex that relies on interactions of the iASPP SILK and RVxF motifs with PP-1c, plus interactions of the PP-1c PxxPxR motif with the iASPP SH3 domain. Small angle X-ray scattering analyses suggest the crystal structure undergoes slow interconversion with more extended conformations in solution. We show that iASPP, and the tumour suppressor ASPP2, enhance the catalytic activity of PP-1c against the small molecule substrate, pNPP as well as p53. The combined results suggest that PxxPxR binding to iASSP SH3 domain is critical for complex formation, and that the modular ASPP-PP-1c interface provides dynamic flexibility that enables functional binding and dephosphorylation of p53 and other diverse protein substrates.

B.2. Introduction

The tumour suppressor p53 is regulated by a complex network of interacting proteins and post-translational modifications (Bieging et al., 2014). A key family of p53 regulatory proteins is the ASPP family (apoptosis-stimulating proteins of p53), which has three members: the highly homologous ASPP1 and ASPP2, and the inhibitory iASPP protein (Trigiante & Lu, 2006). ASPP1 and ASPP2 have been shown to bind the p53 DNA binding domain and potentiate the pro-apoptotic transcriptional activity of p53 (T. Mori et al., 2000; Samuels-Lev et al., 2001; Sgroi et al., 1999). In contrast, iASPP binds to p53 and inhibits pro-apoptotic p53 activity. Given their roles in regulating p53 activity, it is not surprising that ASPP1 or ASPP2 levels are decreased in several human cancers and this is associated with the transition to metastases in breast cancer (Agirre et al., 2006; S. Li et al., 2012; W. K. Liu et al., 2010; Z. J. Liu et al., 2004, 2005; Lossos et al., 2002; S. Mori et al., 2004; T. Mori et al., 2000; Samuels-Lev et al., 2001; Sgroi et al., 1999; Zhao et al., 2010). The iASPP protein is overexpressed in multiple cancers, typically in those that express wild-type p53 (Bergamaschi et al., 2003; Cao et al., 2013; G. Li et al., 2011; S. Li et al., 2012; LIN et al., 2011; W. K. Liu et al., 2010; B. Lu et al., 2010; Pinto et al., 2010; B. Zhang et al., 2011; X. Zhang et al., 2005). In addition, increased ASPP1/2 expression is correlated with enhanced sensitivity of tumour cells to cytotoxic chemotherapy and radiation, while enhanced iASPP expression leads to resistance (Ao et al., 2001; Bergamaschi et al., 2003; Cao et al., 2013; Jia et al., 2014; S. Li et al., 2012; H. Liu et al., 2009; T. Mori et al., 2000; L. Wang et al., 2012; X. Zhang et al., 2005).

All three ASPP proteins share a conserved C-terminal domain consisting of a proline rich region, followed by four ankyrin (ANK) repeats and a SH3 domain that binds p53. While the N-terminal regions of ASPP1/2 and iASPP are poorly conserved and are thought to be mainly unstructured, studies have shown that this region is regulated by multiple post-translational modifications and protein-protein interactions (Aylon et al., 2010; Godin-Heymann et al., 2013; M. Lu et al., 2013; Trigiante & Lu, 2006). The ANK/SH3 domain of ASPP1/2 binds to the DNA binding domain (DBD) of p53 (Gorina & Pavletich, 1996; Patel et al., 2008) and the majority of cancer-associated p53 mutations that inhibit the ability of p53 to bind to DNA also block its interaction with ASPP1/2 (Samuels-Lev et al., 2001). While iASPP has a similar ANK/SH3 domain, it is not thought to bind tightly to the p53 DBD. Instead, proline-rich sequences in other regions of p53 are thought to bind to the iASPP SH3 domain (Ahn et al., 2009; Bergamaschi et al., 2006). Not only is the ANK/SH3 region critical for interactions with p53, but it has also been shown to interact with the serine/threonine protein phosphatase-1c (PP-1c) (Helps et al., 1995; Llanos et al., 2011; Skene-Arnold et al., 2013), fueling speculation that the ASPP proteins may regulate the phosphorylation status of p53.

Together with Protein Phosphatase-2A (PP-2A), PP-1c handles most Ser/Thr dephosphorylation in human cells (Shi, 2009) and gains its specificity through selective interactions with hundreds of regulatory proteins that target PP-1c to specific phosphorylated substrates (Bollen et al., 2010; Brautigan & Shenolikar, 2018). A growing body of structural studies have indicated some conserved themes in PP-1c regulatory proteins. Many of these proteins are largely unstructured, and utilize extended coil structures to wrap around PP-1c to contact various targeting surfaces (Bollen et al., 2010; Choy et al., 2012). Not only are these interactions thought to bind the regulatory protein to PP-1c, but they can also block potential substrate binding grooves on the PP-1c surface to modulate substrate specificity. The most common PP-1c binding motif identified is the tetrapeptide RVxF motif ([R/K]-[V/I/L]-X-[F/W], where X can be any amino acid except proline), and other secondary motifs have been identified such as the MyPhone, SILK ([S/G]-I-L-K), and ΦΦ motifs (Bollen et al., 2010; Choy et al., 2012). In addition to these short peptide motifs, ANK domains have also been shown to bind PP-1c, as in the PP-1c regulatory factor, MYPT1 (Terrak et al., 2004). MYPT1 also binds to PP-1c via MyPhone and RVXF motifs N-terminal to its ANK repeats.

The fact that all three ASPP proteins contain RVxF-like motifs immediately N-terminal to their ANK domains suggested that they might bind PP-1c in a manner similar to MYPT. Building on this work, we recently showed that purified ASPP1, ASPP2 and iASPP can all form 1:1 stoichiometric complexes with PP-1c that critically involve the unstructured C-terminal tail of PP-1c (Skene-Arnold et al., 2013). Here we determine the crystal structure of iASPP bound to PP-1c, which reveals that the interactions between the two proteins are stabilized by several distinct contact surfaces. Small angle X-ray scattering however indicates that the complex accesses additional conformational states in solution that likely involve the release of one or more of the peptide contacts. From the combined results, we propose a structure-based model for how the modular nature of the ASPP-PP-1c contact surface provides dynamic flexibility that enables p53 binding and dephosphorylation.

B.3. Results

B.3.1. Overview of the crystal structure of iASPP$_{608\text{-}828}$ bound to PP-1ca

To understand the structural basis for interactions between iASPP and PP-1c, we crystallized and determined the structure of the minimal PP-1c-binding iASPP fragment, iASPP$_{608\text{-}828}$ (Skene-Arnold et al., 2013), bound to PP-1cα at 3.41 Å resolution (Figure B-1, Table B-1, see the section B.5. Materials and Methods for more details). iASPP residues 627-828 constitute the folded core of iASPP, and is composed of four N-terminal ankyrin (ANK) repeats and a C-terminal SH3 domain that we find are essentially identical to the structures of this region in free iASPP (Robinson et al., 2008) as well as ASPP2 (Gorina & Pavletich, 1996). The iASPP N-terminal ANK repeat docks against a shallow groove on PP-1c and the N-terminal peptide iASPP$_{608\text{-}626}$ extends away from this contact surface, making two additional interactions with the PP-1c catalytic core (Figure B-1A, B). The first contact involves residues $_{622}$RARL$_{625}$, which plays the role of an RVxF motif (Figure B-2A) as previously predicted (Skene-Arnold et al., 2013). N-terminal to this sequence the electron density becomes poor and tracks away from PP-1c, however, additional

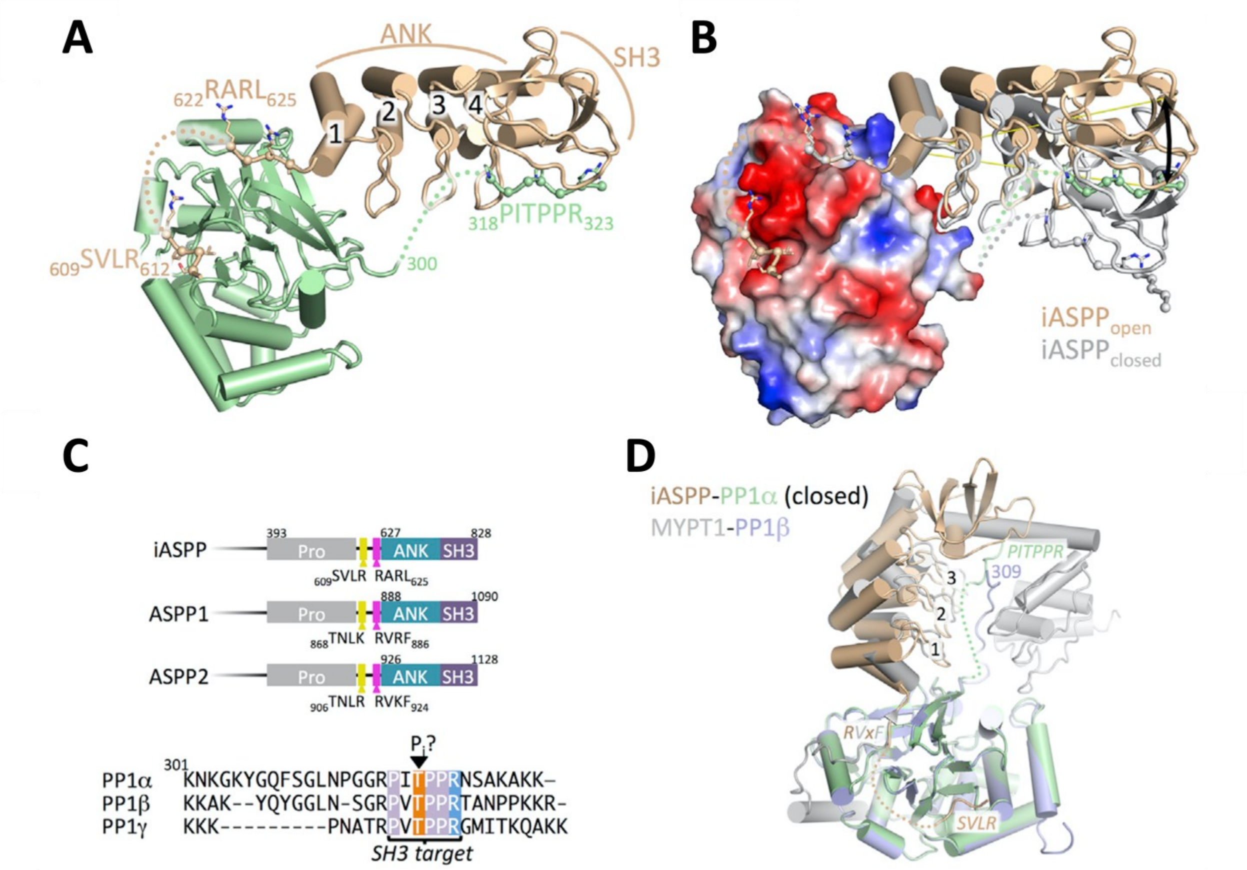
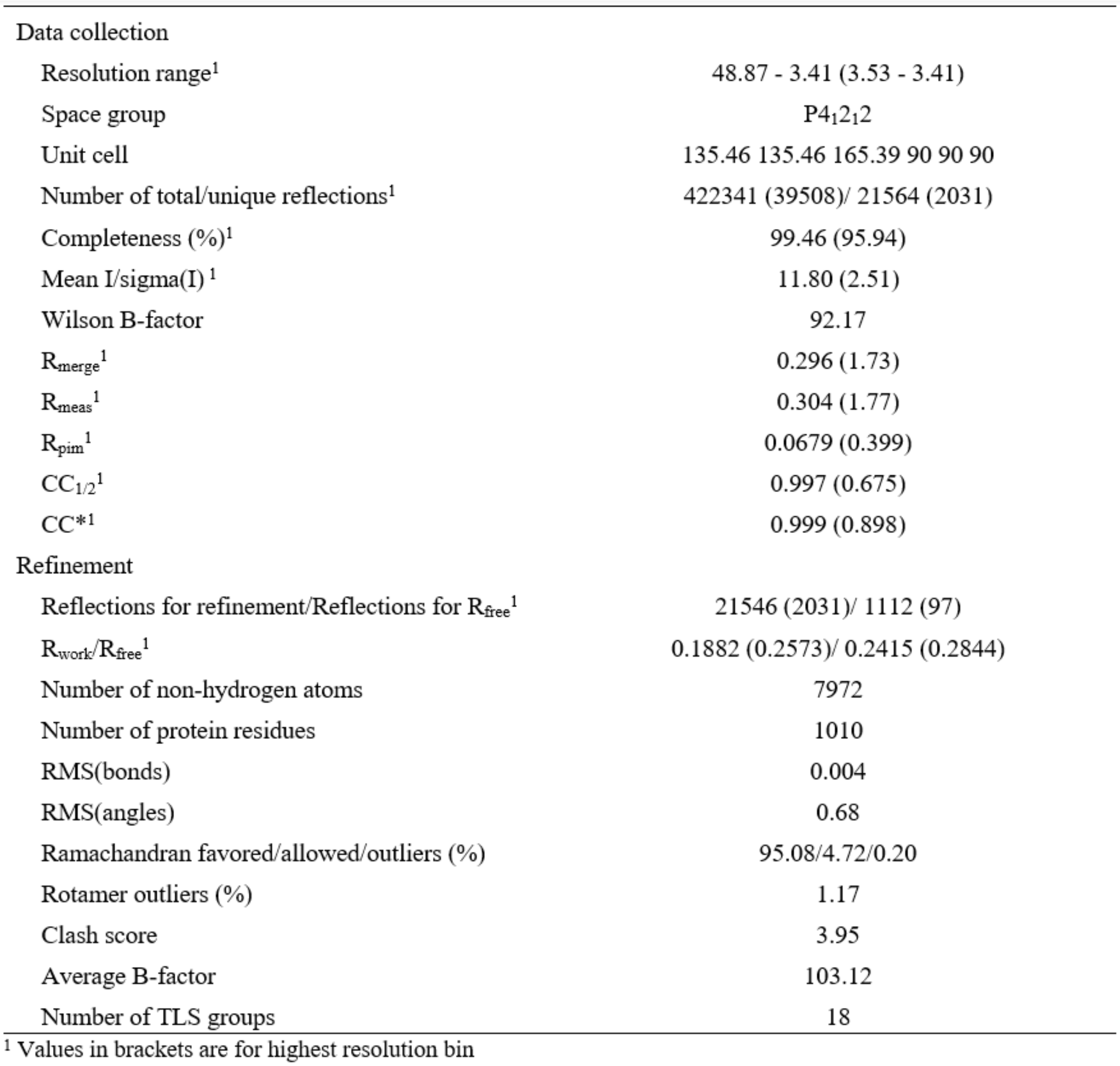


Figure B-1. Crystal structure of iASPP$_{608\text{-}828}$ bound to PP-1cα.

(A) Overview of the structure of iASPP$_{608\text{-}828}$ bound to PP-1cα. iASPP is brown, PP-1cα is in green. Peptide segments involved in key protein-protein contacts are illustrated with Cα spheres and the four ANK repeats are numbered. Regions of polypeptide chain not visible in the electron density are indicated by dotted lines. (B) Comparison of the two iASPP$_{608\text{-}828}$-PP-1cα complexes in the asymmetric unit, aligned on PP-1cα, with the second iASPP in the closed conformation colored grey and an electrostatic charge surface displayed for PP-1cα. The ~22° angular difference in the orientation of the two iASPPs relative to PP-1cα is indicated. (C) Top panel, comparison of the primary structures of the C-terminal regions of iASPP, ASPP1 and ASPP2 with key domains and motifs indicated. Bottom panel, sequence alignment of the C-terminal unstructured tails of the three PP-1cα isoforms. (D) Structural comparison of the closed complex of iASPP-PP-1cα with the structure of MYPT1-PP-1β, aligned on PP-1cα with the different chains colored as indicated in the figure. Crystallographic statistics are given in Table B-1.



Data collection	
Resolution range[1]	48.87 - 3.41 (3.53 - 3.41)
Space group	$P4_12_12$
Unit cell	135.46 135.46 165.39 90 90 90
Number of total/unique reflections[1]	422341 (39508)/ 21564 (2031)
Completeness (%)[1]	99.46 (95.94)
Mean I/sigma(I)[1]	11.80 (2.51)
Wilson B-factor	92.17
R_{merge}[1]	0.296 (1.73)
R_{meas}[1]	0.304 (1.77)
R_{pim}[1]	0.0679 (0.399)
$CC_{1/2}$[1]	0.997 (0.675)
CC*[1]	0.999 (0.898)
Refinement	
Reflections for refinement/Reflections for R_{free}[1]	21546 (2031)/ 1112 (97)
R_{work}/R_{free}[1]	0.1882 (0.2573)/ 0.2415 (0.2844)
Number of non-hydrogen atoms	7972
Number of protein residues	1010
RMS(bonds)	0.004
RMS(angles)	0.68
Ramachandran favored/allowed/outliers (%)	95.08/4.72/0.20
Rotamer outliers (%)	1.17
Clash score	3.95
Average B-factor	103.12
Number of TLS groups	18

[1] Values in brackets are for highest resolution bin

Table B-1. Data collection and refinement statistics.

density is observed packed into an adjacent groove on PP-1c groove that can bind SILK-like motifs. We suggest this density represents the N-terminal-most region, $_{609}$SVLR$_{612}$, which may play the role of a SILK motif (Figure B-2B). Notably, the conservation of the $_{609}$SVLR$_{612}$ and $_{622}$RARL$_{625}$ sequences in the ASPP protein family suggest that similar interactions may stabilize complexes with PP-1c in the other ASPP proteins (Figure B-1C). The final major contact between PP-1c and iASPP involves the SH3 domain of iASPP and a region of left-handed helical density that we have modeled as the $_{318}$PxxPxR$_{323}$ motif near the C-terminus of the flexible PP-1c tail

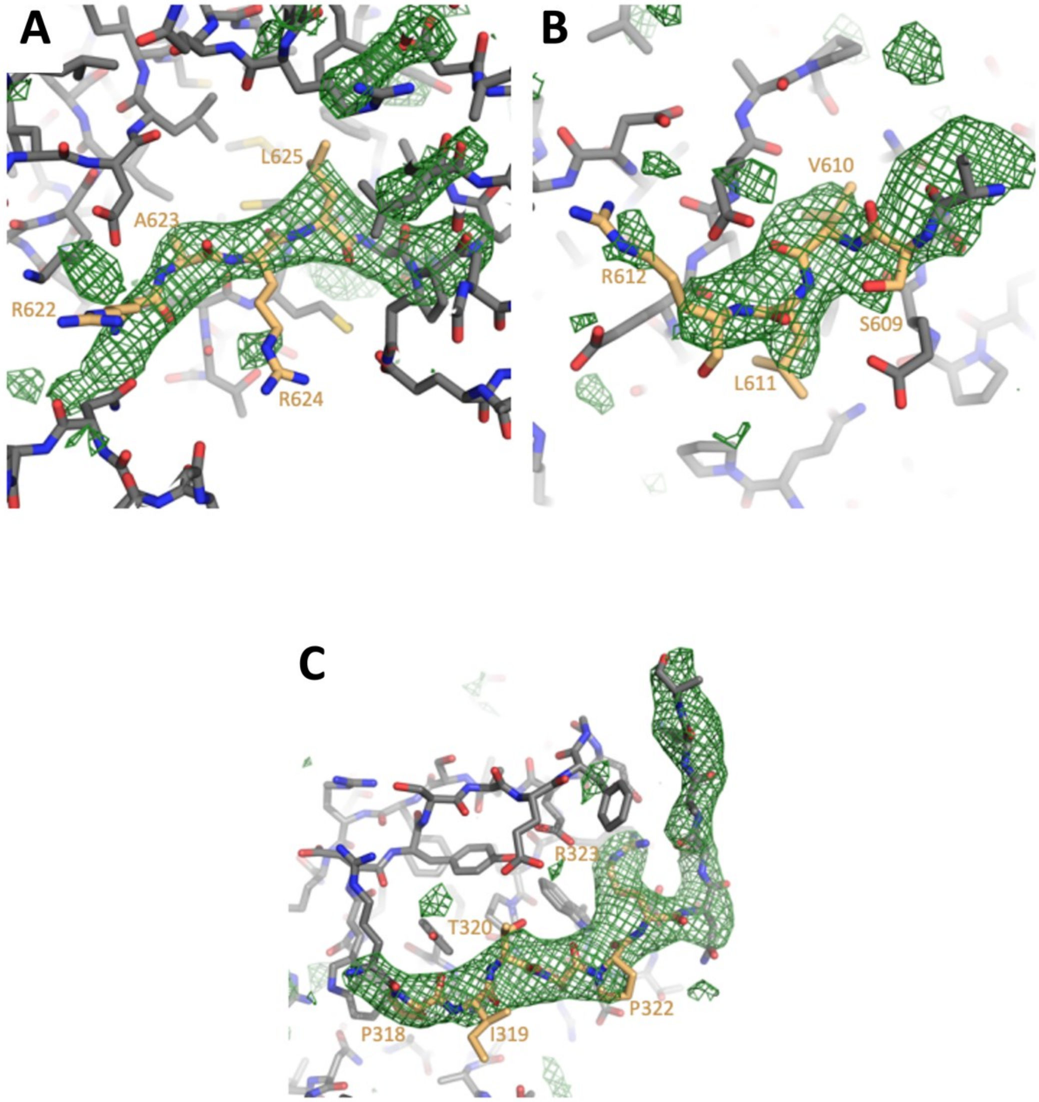


Figure B-2. iASPP-PP-1cα representative electron density.

Related to Figures B-1 and B-3. Shown is representative F_o-F_c difference electron density (green), phased with $\text{iASPP}_{627\text{-}828}$ and $\text{PP-1c}\alpha_{1\text{-}300}$ placed by molecular replacement before refinement, and contoured at 2.5 σ. Difference density is shown for the RARL (A), SILK (B) and PxxPxR (C) motifs. In each panel, the omitted motif is shown as beige sticks and residues are labelled.

(Figure B-2C). This motif is conserved in all PP-1c isoforms, although the length of the tail linking the PxxPxR motif to the catalytic domain varies between isoforms (Figure B-1C).

Two distinct iASPP-PP-1c complexes are observed in the crystallographic asymmetric unit. The overall structures are similar, however a structural alignment of the PP-1c catalytic domains of two complexes (Figure B-1B) reveals conformational differences between the binding orientation of the iASPP ANK-SH3 domain relative to PP-1c. The difference between the two complexes likely reflects the conformational flexibility of the complex in solution, as the alignment shows that iASPP pivots on the N-terminal ANK repeat such that there is an ~20o difference in the angular orientation of iASPP with respect to PP-1c between the two complexes. The MYPT family of PP-1c regulatory subunits also use an RVxF motif N-terminal to ANK repeats to bind PP-1c (Grassie et al., 2011; Ito et al., 2004) and the overall orientation of the more compact or closed iASPP-PP-1c complex resembles the PP-1c binding orientation of the MYPT1 regulatory subunit (Figure B-1D) (Terrak et al., 2004).

B.3.2. Four distinct contact surfaces stabilize the iASPP-PP-1c complex

Our structure shows that a large ~3100 Å^2 solvent accessible surface area is buried upon formation of the iASPP-PP-1c complex, explaining the tight binding affinity of the complex (26 nM (Skene-Arnold et al., 2013)). The two proteins do not interact through a single contiguous binding surface; however, they instead contact one another through four distinct interfaces (Figure B-3A). The largest of these interfaces is formed between the PxxPxR motif of PP-1c and the SH3 domain of iASPP, which buries ~990 Å^2 of the solvent accessible surface (Figure B-3B and Figure B-2C). The PP-1c C-terminal PxxPxR motif adopts a polyproline helical structure similar to the classical interactions observed between SH3 domains and class II polyproline peptide targets (Feng et al., 1994; Lim et al., 1994). Within the iASPP SH3 domain, Trp767, Tyr814, Pro811 and Trp798 form the base of the peptide binding cleft and contact PP-1c residues Pro318 and Pro321, which protrude from the PxxPxR motif. Arg323 of PP-1c packs against a cluster of negatively charged residues from the iASPP SH3 RT loop (Glu772, Asp775, Glu776). The importance of the PxxPxR-SH3 contact is consistent with the finding that deletion of the PP-1c tail abrogates interactions with ASPP proteins in pull-down experiments (Skene-Arnold et al., 2013). The PxxPxR motif also contains a conserved Thr320 (Figure B-1C), which packs into a pocket formed by Tyr769, Trp798, Pro811, Tyr814, and Glu772 (Figure B-3B). Thr320 is a known cdk2-cyclin A phosphorylation site that is phosphorylated during mitosis and is associated with PP-1c auto-inhibition (Berndt et al., 1997; Dohadwala et al., 1994). Based on our iASPP-PP-1c crystal structure, phosphorylation of PP-1c Thr320 would sterically interfere with the packing of this residue into the ASPP SH3

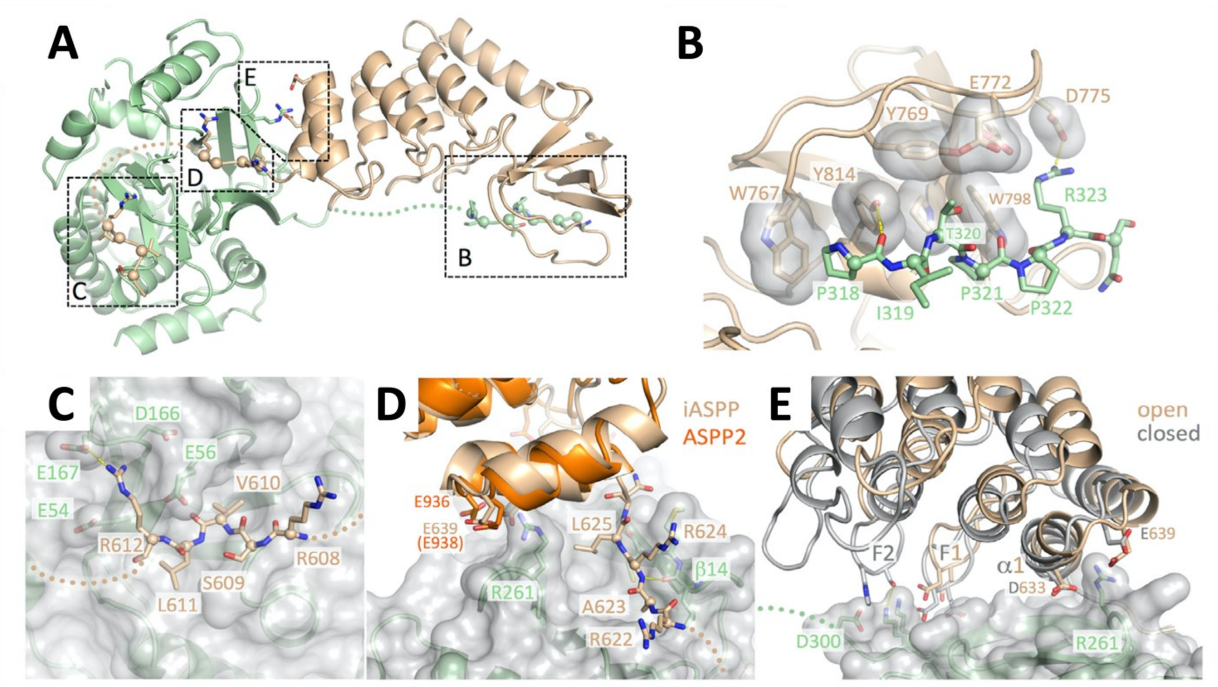


Figure B-3. Four distinct contact surfaces stabilize the iASPP-PP-1c complex.

(A) Overview of the iASPP$_{608\text{-}828}$-PP-1cα structure with the four intermolecular contact surfaces boxed and labeled corresponding to the subsequent panels. (B) Details of the SH3-PxxPxR motif interaction. Key residues from the iASPP SH3 domain that interact with the PxxPxR motif are shown as sticks with transparent surfaces. The PxxPxR peptide is shown in sticks with Cα spheres displayed. (C) Details of the SILK-PP-1c interaction. A transparent surface is shown for PP-1c with key side chains shown as sticks. The SILK peptide is shown in sticks with Cα spheres displayed. (D) Details of the RARL-PP-1c interaction and ANK-PP-1c contact. The RARL peptide is shown in sticks with Cα spheres displayed, while PP-1c is displayed with a transparent surface and key contact side chains as sticks. The ASPP2 structure (PDB ID: 1YCS) is shown aligned on the iASPP structure with conserved acidic residues in the first ANK repeat shown as sticks. (E) Interactions between the iASPP ANK repeats and PP-1c. Shown are the two iASPP-PP-1c complexes aligned on PP-1c. The fingers from the first two ankyrin repeats are labeled F1 and F2 and residues that make intermolecular contacts are shown as sticks. Electron density for each of these contact surfaces is shown in Figure B-2.

binding pocket and would also introduce electrostatic repulsions with iASPP Glu772. Indeed, we have previously shown that a phospho-mimetic mutant of PP-1c (PP-1cα T311D) destabilizes PP-1c-ASPP interactions (Skene-Arnold et al., 2013).

The second largest contact surface involves the iASPP $_{609}$SVLR$_{612}$ motif that packs into a depression on the PP-1c catalytic domain and buries ~790 Å^2 of solvent accessible surface area (Figure B-3C and Figure B-2B). This interaction is essentially identical to the interaction of the "SILK" motif of inhibitor 2 with PP-1c (Hurley et al., 2007). Val610 and Leu611 pack into hydrophobic pockets in PP-1c while Arg612 makes electrostatic contact with an extensive negatively charged PP-1c surface composed of Glu167, Asp166, Glu56, and Glu54.

The next largest contact surface involves iASPP RVxF-like motif ($_{622}$RARL$_{625}$), which buries ~600 Å^2 of solvent accessible surface area (Figure B-3D and Figure B-2C). As observed in other RVXF motif interactions (Bollen et al., 2010), the second and fourth residues of the motif (Ala623 and Leu625) form a key part of the interface by binding to hydrophobic pockets on PP-1c. Further anchoring the interaction are main chain hydrogen bonds between the iASPP RARL motif and PP-1c β14. Mutation of Leu625 to Ala significantly weakens the interaction with PP-1c in pull-down experiments (Skene-Arnold et al., 2013). In addition, deletion of both the SILK and RARL motifs (iASPP$_{626-828}$), decreases but does not completely abolish iASPP-PP-1c interactions, suggesting that these contact sites may not be as important as the PxxPxR-SH3 interaction (Skene-Arnold et al., 2013).

The final point of interaction is between the iASPP ANK repeats and PP-1c catalytic domain (Figure B-3D and E). The interaction is not as extensive as the other contact surfaces and only buries a maximum of ~340 Å^2 solvent accessible surface area in the closed complex, less in the more open form. The interaction pivots on α1 of the first ANK motif and involves electrostatic interactions between PP-1c Arg261 and a conserved negatively charged surface encompassing the C-terminus of α 1 and N-terminus of α 2 (Figure B-3D and E). The pivoting of iASPP allows additional contacts between the fingers of the first two iASPP ANK repeats (F1 and F2) and the PP-1c catalytic domains (Figure B-3E). These contacts likely limit the degree of closure of the complex. While the surface area buried in this interaction is modest, mutation of Arg261 to Ser has been shown to significantly reduce ASPP-PP-1c affinity in pull-down assays (Skene-Arnold et al., 2013).

B.3.3. Flexibility of ASPP-PP-1c complexes revealed by SEC-SAXS

To characterize the structure and flexibility of ASPP-PP-1c complexes in solution by SAXS (Rambo & Tainer, 2013), we applied SAXS to ASPP-PP-1c complexes separated by in-line size exclusion chromatography (SEC-SAXS; Figure B-4). In this method, purified iASPP$_{608\text{-}828}$-PP-1cα or ASPP2$_{905\text{-}1128}$-PP-1cα was run on size exclusion chromatography in-line with SAXS and multi-angle laser light scattering (MALS). For both samples, the complexes eluted as a single asymmetric peak (Figure B-4A and Figure B-5). MALS analysis of this peak suggested a small amount of aggregation near the leading edge of the peak, however the mass across the majority of the peak corresponded closely to the expected mass of the heterodimer (ASPP2$_{905\text{-}1128}$-PP-1ca [MW] [theoretical] = 62.4 kDa; MW [MALLS] = 69.6 ± 0.4 kDa). Automated Guinier calculation of the radius of gyration (RG) across the peak revealed a significant decrease in the R_G from ~35 Å at the peak, to ~25-30 Å in the tail (Figure B-4A). Taken together, these data suggest that heterodimeric iASPP-PP-1c adopts a mixture of different conformational states that slowly interconvert over the time course of the chromatography. Similar results were obtained with ASPP2-PP-1c complexes (Figure B-5).

To further analyze the structure and dynamics of these conformational states, we used the program RAW (Hopkins et al., 2017) to deconvolute the scattering data into two separate sets representing compact (R_G = 31.7 Å; D_{max} = 113 Å) and extended (R_G = 34.5 Å; D_{max} = 121 Å) conformational states (Figure B-5). To address whether this scattering data could be explained by the iASPP-PP-1c crystal structure, we used a minimal ensemble search (MES) of diverse conformational libraries of ASPP-PP-1c ((Pelikan et al., 2009); see Table B-2 details of libraries used and the MES procedure). The compact data subset could be fit with a model derived from the crystal structure (χ^2 = 3.2; library "crystal", Figure B-4B, C, Table B-2), and the quality of the fit could be improved by allowing flexibility in the docking of the ASPP ANK/SH3 domain against PP-1c (χ^2 = 1.7; library 5001, Figure B-4C, Table B-2). The extended data set, however, was only poorly approximated by models from these libraries (Figure B-4D, Table B-2), suggesting that more extended models would be required to account for this scattering. To allow further flexibility, we generated model libraries in which one or more of the RVxF, SILK or PxxPxR peptides were released from their respective binding partners. MES analysis using these model libraries for the compact data subsets gave ensembles that contained models resembling the crystal structure as

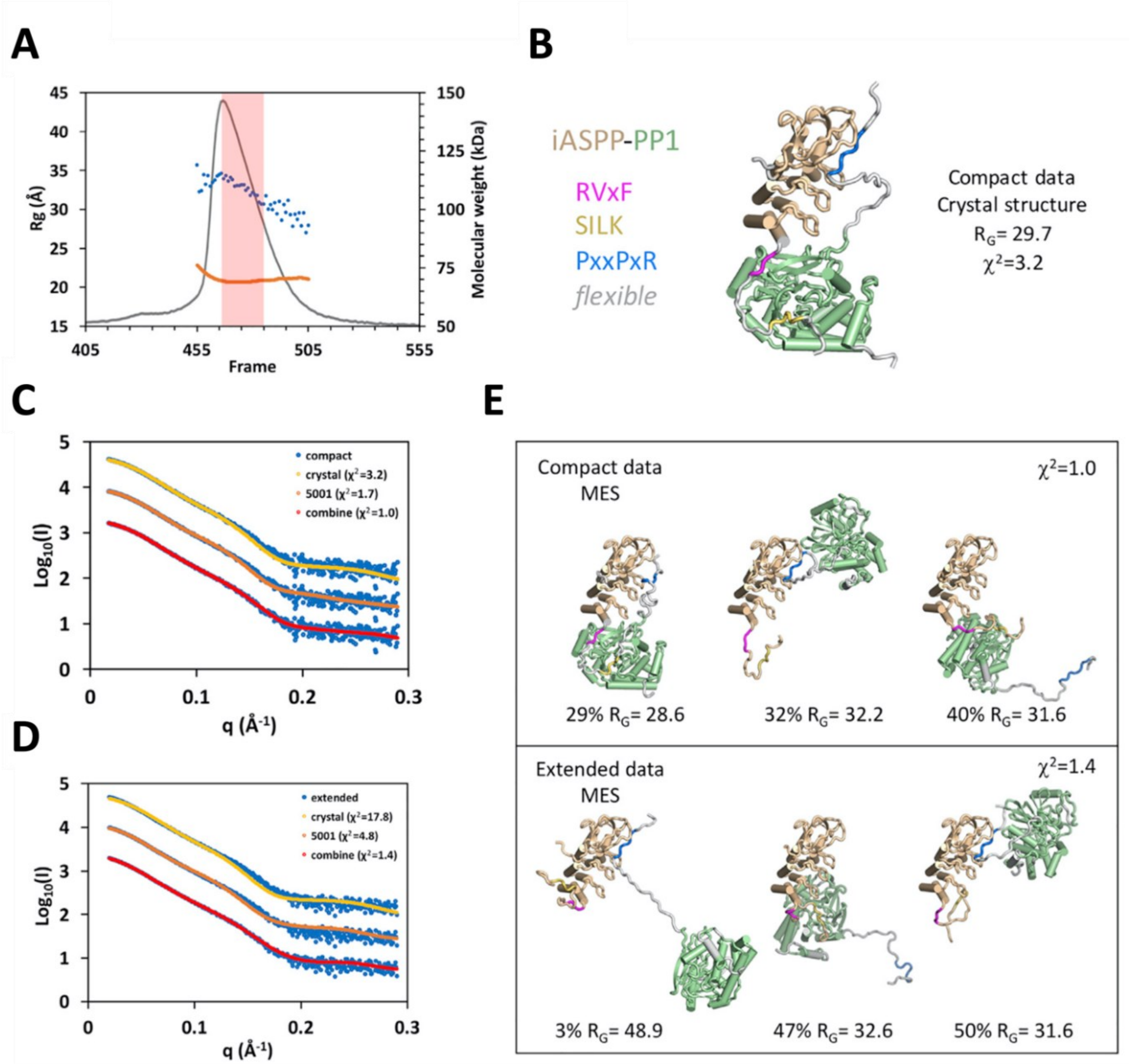


Figure B-4. Conformational flexibility of iASPP-PP-1c revealed by SAXS.

(A) SEC-SAXS chromatogram with raw scattering trace in grey, the blue dots give the R_G values calculated by Guinier approximation for each frame, and the orange trace gives the MW as determined by MALS. (B) MES modeling using the 5011 library (based on the open form of the iASPP-PP-1c crystal structure). The single model is shown with color coding as shown on the left. (C) and (D) Fit of the compact (C) and extended (D) data sets to calculated scattering from the MES models derived from the 5011 library (crystal), the 5001 library, and the combined libraries. (E) Three model ensembles derived from the combined libraries for the compact data set (top panel) and the extended data set (bottom panel). Additional details and a comparable analysis of ASPP2-PP-1c and iASPP(621-828)-PP-1c are given in Figures B-5, 6, 7 and 11 and Tables B-1 and 2.

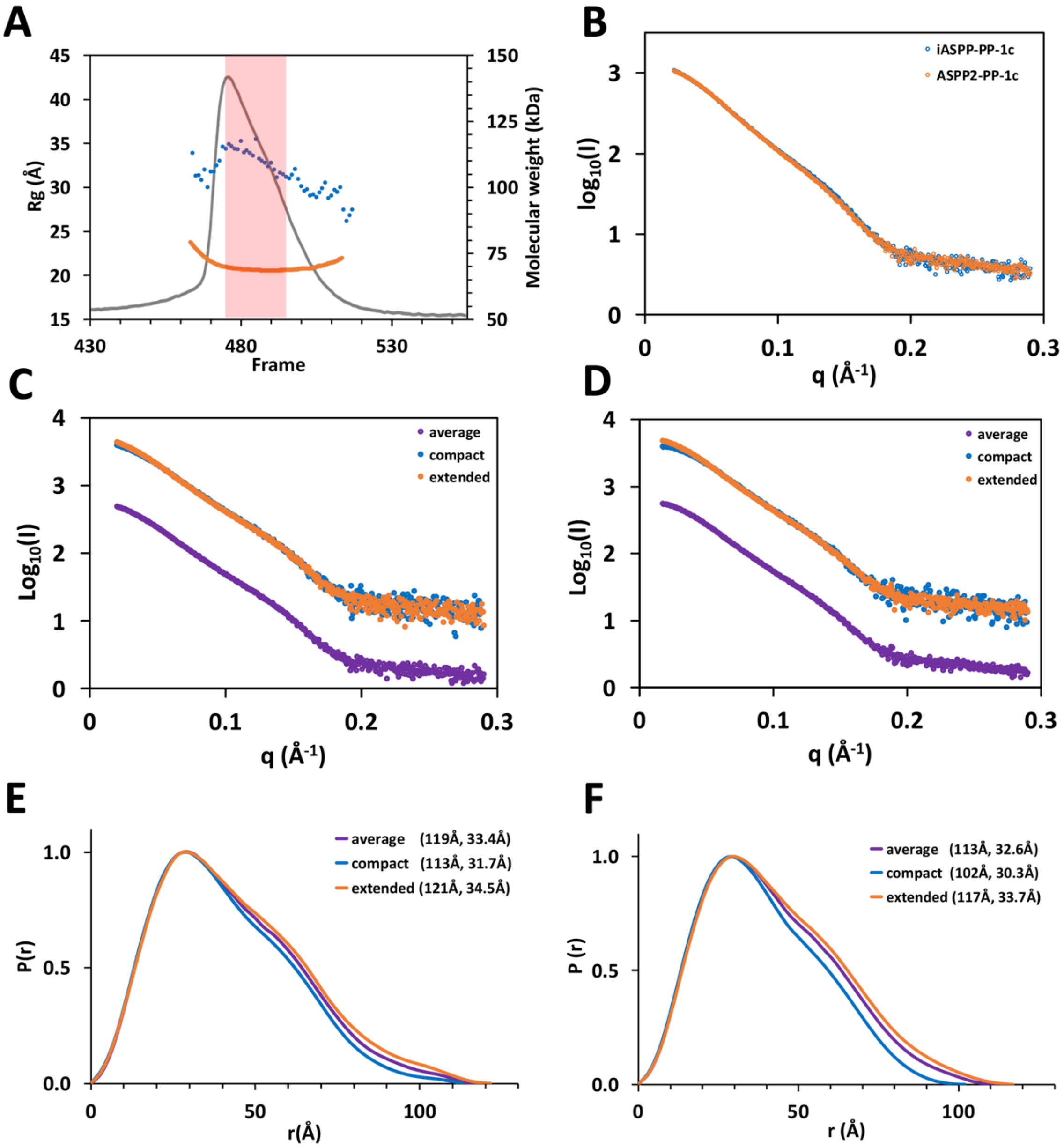


Figure B-5. SEC-SAXS analysis of iASPP-PP-1c and ASPP2-PP-1c.

Related to Figure B-4. (A) SEC-SAXS on ASPP2-PP-1c. The grey trace corresponds to the size exclusion UV$_{280}$ trace, the blue dots give the R$_G$ values calculated by Guinier approximation for each frame, and the orange trace gives the MW as determined by MALLS. (B) Comparison of the scattering from iASPP-PP-1c with ASPP2-PP-1c, integrated from peak maximum to ½ maximal peak height (pink region in (A)). (C) and (D) RAW deconvolution of scattering into compact and extended subsets for iASPP-PP-1c (C) and ASPP2-PP-1c (D). (E) and (F) P(r) plots for RAW deconvoluted scattering for iASPP-PP-1c (E) and ASPP2-PP-1c (F).

Table B-2. MES analysis of iASPP-PP-1c SAXS data.

Pool (library)[1]		Experimental Data: iASPP-PP-1c		
		Deconvoluted Small[2]	**Deconvoluted Large[2]**	**Non-deconvoluted[2]**
5011	R_G Range of Pool	28.9 - 29.7		
	D_{max} Range of Pool	90 - 110		
	FOXS Chi^2 Range of Pool[3]	3.2 – 10.7	17.8 – 41.2	25.3 – 69.3
	Chi^2 of best model/ensemble	3.2	17.8	25.3
	Pool ID/R_G/D_{max}/%[4]	5011 29.7 97.2 100	5011 29.7 97.2 100	5011 29.7 97.2 100
5013	R_G Range of Pool	27.4 - 28.3		
	D_{max} Range of Pool	84 - 100		
	FOXS Chi^2 Range of Pool	9.3 – 23.2	34.0 – 70.1	57.9 – 129.3
	Chi^2 of best model/ensemble	9.3	34.0	57.9
	Pool ID/R_G/D_{max}/%	5013 28.3 88.5 100	5013 28.3 88.5 100	5013 28.3 88.5 100
Combined Pools 5001, 5002, 5003, 5004, 5005, 5006, 5007	Chi^2 of 3 model ensemble	1.0	1.4	1.4
	Pool ID/R_G/D_{max}/%	5005 31.6 120 40	5007 31.6 105 50	5007 32.1 112 44
		5007 32.2 113 32	5005 32.6 125 47	5006 34.7 144 34
		5006 28.6 89.7 29	5007 48.9 130 3	5007 27.8 95.4 22

---

[1] Identifier for the model library (pool) created using BILBO-MD

[2] Experimental SAXS data was deconvoluted using RAW

[3] Chi^2 are calculated for the individual models in the pool

[4] Pool ID: the code representing the BILBO-MD library; R_G: radius of gyration for the selected model; Dmax: maximum pairwise interatomic distance in the selected model; %: fraction of the ensemble represented by the model

Pool (library)		Experimental Data: iASPP-PP-1c		
		Deconvoluted Small	**Deconvoluted Large**	**Non-deconvoluted**
5001	R_G Range of Pool	28.1 - 33.6		
	D_{max} Range of Pool	89 - 126		
	FOXS Chi^2 Range of Pool	1.8 - 16	5.6 – 54	8.0 - 95
	Chi^2 of best model/ensemble	1.7	4.8	6.7
	Pool ID/R_G/Dmax/%	5001 30.2 98.1 74	5001 32.4 115 53	5001 30.4 95.0 44
		5001 32.2 120 20	5001 32.2 120 25	5001 32.2 120 40
		5001 28.5 89.5 6	5001 30.4 95.0 23	5001 32.4 115 15
5002	R_G Range of Pool	27.9 - 37.3		
	D_{max} Range of Pool	89 - 185		
	FOXS Chi^2 Range of Pool	2.3 – 31	4.4 - 80	6.0 - 149
	Chi^2 of best model/ensemble	1.1	2.7	2.9
	Pool ID/R_G/Dmax/%	5002 28.9 94.2 54	5002 34.1 144 46	5002 34.1 144 41
		5002 31.3 97.5 24	5002 31.4 100 36	5002 31.3 98.9 30
		5002 34.8 138 22	5002 28.8 92.1 17	5002 28.9 94.2 29
5003	R_G Range of Pool	28.1 - 34.7		
	D_{max} Range of Pool	89 - 152		
	FOXS Chi^2 Range of Pool	1.2 – 33	3.6 - 68	4.4 - 122
	Chi^2 of best model/ensemble	1.1	2.3	2.4
	Pool ID/R_G/Dmax/%	5003 28.9 91.9 45	5003 33.8 138.9 49	5003 33.8 139 44
		5003 33.0 134 32	5003 31.2 100 34	5003 28.9 96 35
		5003 31.2 99.8 24	5003 28.9 95.7 18	5003 31.2 100 21
5004	R_G Range of Pool	25.8 - 46.4		
	D_{max} Range of Pool	75 - 139		
	FOXS Chi^2 Range of Pool	2.0 - 187	3.3 - 320	3.9 - 721
	Chi^2 of best model/ensemble	1.0	2.3	2.3

Pool (library)		Experimental Data: iASPP-PP-1c											
	Pool ID/R_G/Dmax/%	5004	31.7	111	65	5004	31.8	108	46	5004	31.0	91.5	66
		5004	27.6	81.9	32	5004	31.0	91.5	45	5004	40.6	128	17
		5004	43.5	126	3	5004	43.4	124	9	5004	27.4	9`1.4	17
5005	R_G Range of Pool	26.1 - 54.3											
	D_{max} Range of Pool	73 - 225											
	FOXS Chi^2 Range of Pool	1.3 - 306				1.9 - 518				2.7 - 1162			
	Chi^2 of best model/ensemble	1.0				1.7				1.6			
	Pool ID/R_G/Dmax/%	5005	32.3	117	35	5005	32.3	117	37	5005	32.3	117	59
		5005	31.3	120	35	5005	32.6	129	36	5005	32.7	128	29
		5005	28.9	94.2	29	5005	32.0	125	27	5005	26.7	89	12
5006	R_G Range of Pool	27.7 – 38.0											
	D_{max} Range of Pool	89.1 - 208											
	FOXS Chi^2 Range of Pool	1.9 - 57				2.9 - 94				3.0 - 178			
	Chi^2 of best model/ensemble	1.1				2.0				2.0			
	Pool ID/R_G/Dmax/%	5006	28.6	90.2	44	5006	34.6	134	53	5006	34.6	134	45
		5006	31.5	98.3	34	5006	28.8	91.9	24	5006	28.8	91.9	34
		5006	34.3	132	22	5006	31.8	98.3	23	5006	31.8	98.3	21
5007	R_G Range of Pool	24.3 - 56.1											
	D_{max} Range of Pool	71 - 202											
	FOXS Chi^2 Range of Pool	1.4 - 378				2.5 - 676				5.0 - 1525			
	Chi^2 of best model/ensemble	1.0				1.3				1.3			
	Pool ID/R_G/Dmax/%	5007	31.6	106	50	5007	31.6	105	62	5007	31.6	105	44
		5007	28.6	91	41	5007	38.1	121	22	5007	28.1	90.3	29
		5007	37.8	118	9	5007	28.0	92.9	16	5007	36.9	135	27

Table B-3. MES analysis of ASPP2-PP-1c SAXS data.

Pool (library)[1]		Experimental Data: ASPP2-PP-1c		
		Deconvoluted Small[2]	Deconvoluted Large[2]	Non-deconvoluted[2]
5011	R_G Range of Pool	28.9 - 29.7		
	D_{max} Range of Pool	90 - 110		
	FOXS Chi^2 Range of Pool[3]	1.3 – 4.8	22 – 56	21 – 72
	Chi^2 of best model/ensemble	1.3	22	21
	Pool ID/R_G/Dmax/%[4]	5011 29.5 94.2 66	5011 29.7 97.2 100	5011 29.7 97.2 100
		5011 29.6 105 25		
		5011 29.7 98.0 9		
5013	R_G Range of Pool	27.4 - 28.3		
	D_{max} Range of Pool	84 - 100		
	FOXS Chi^2 Range of Pool	3.9 – 12	45 – 99	57 - 143
	Chi^2 of best model/ensemble	3.9	45	57
	Pool ID/R_G/Dmax/%	5013 28.3 88.5 100	5013 28.3 88.5 100	5013 28.3 88.5 100
Combined Pools 5001, 5002, 5003, 5004, 5005, 5006, 5007	Chi^2 of 3 model ensemble	1.2	1.3	1.3
	Pool ID/R_G/Dmax/%	5007 29.1 89.7 36	5005 32.4 120 57	5007 32.1 115 53
		5005 31.3 109 36	5007 31.6 105 40	5005 32.2 122 25
		5001 28.9 90.0 28	5007 24.6 71.8 2	5003 28.6 94.9 22

---

[1] Identifier for the model library (pool) created using BILBO-MD

[2] Experimental SAXS data was deconvoluted using RAW

[3] Chi2 are calculated for the individual models in the pool

[4] Pool ID: the code representing the BILBO-MD library; R_G: radius of gyration for the selected model; Dmax: maximum pairwise interatomic distance in the selected model; %: fraction of the ensemble represented by the model

Pool (library)		Experimental Data: ASPP2-PP-1c		
		Deconvoluted Small	Deconvoluted Large	Non-deconvoluted
5001	R_G Range of Pool	28.3 - 33.6		
	D_{max} Range of Pool	89.2 - 126		
	FOXS Chi^2 Range of Pool	2.3 – 22	5.8 - 76	7.3 - 104
	Chi^2 of best model/ensemble	1.6	4.9	5.7
	Pool ID/R_G/Dmax/%	5001 30.2 98.1 58	5001 32.4 115 46	5001 32.2 120 44
		5001 29.0 91.6 38	5001 30.4 95.0 29	5001 30.4 95.0 40
		5001 28.4 91.7 4	5001 32.2 120 25	5001 30.2 98.1 16
5002	R_G Range of Pool	27.9 - 37.3		
	D_{max} Range of Pool	89 - 185		
	FOXS Chi^2 Range of Pool	1.7 – 38	4.4 - 114	6.9 - 169
	Chi^2 of best model/ensemble	1.2	2.8	2.8
	Pool ID/R_G/Dmax/%	5002 28.7 91.3 42	5002 31.8 105 46	5002 28.9 94.2 44
		5002 31.3 98.4 37	5002 34.3 144 29	5002 32.0 103 38
		5002 28.6 91.5 21	5002 28.8 92.1 25	5002 34.1 144 18
5003	R_G Range of Pool	28.1 - 34.7		
	D_{max} Range of Pool	89.1 - 152		
	FOXS Chi^2 Range of Pool	1.4 – 43	3.4 – 97	3.2 - 138
	Chi^2 of best model/ensemble	1.2	1.9	1.7
	Pool ID/R_G/Dmax/%	5003 30.8 116.3 44	5003 33.8 119 42	5003 28.9 93.4 41
		5003 28.9 90.5 33	5003 31.2 100 40	5003 33.8 119 37
		5003 28.9 94.0 24	5003 28.9 95.7 18	5003 31.2 100 22
5004	R_G Range of Pool	25.8 - 46.4		
	D_{max} Range of Pool	75.4 - 139		
	FOXS Chi^2 Range of Pool	1.8 - 183	2.7 - 511	3.2 - 1029
	Chi^2 of best model/ensemble	1.2	2.2	1.9

Pool (library)		Experimental Data: ASPP2-PP-1c											
	Pool ID/R_G/Dmax/%	5004	31.7	105	41	5004	31.5	105	68	5004	31.5	106	58
		5004	27.2	83.4	31	5004	30.9	88.5	24	5004	34.5	112	22
		5004	28.9	92.0	28	5004	41.9	120	7	5004	27.0	77.0	21
5005	R_G Range of Pool	26.1 - 54.3											
	D_{max} Range of Pool	73.1 - 225											
	FOXS Chi^2 Range of Pool	2.4 - 300				1.8 - 815				2.3 - 1652			
	Chi^2 of best model/ensemble	1.2				1.4				1.5			
	Pool ID/R_G/Dmax/%	5005	29.5	95.3	40	5005	32.3	119	47	5005	32.3	117	52
		5005	28.4	87.7	31	5005	32.3	117	42	5005	31.8	110	28
		5005	31.6	103	28	5005	26.9	90.4	10	5005	26.8	89.1	20
5006	R_G Range of Pool	27.7 – 38.0											
	D_{max} Range of Pool	89.1 - 208											
	FOXS Chi^2 Range of Pool	1.4 – 65				2.6 - 137				3.3 - 249			
	Chi^2 of best model/ensemble	1.2				2.0				1.9			
	Pool ID/R_G/Dmax/%	5006	31.6	114	41	5006	34.5	132	41	5006	31.6	98.4	41
		5006	28.7	91.5	34	5006	31.8	102	36	5006	28.7	94.7	31
		5006	28.6	89.7	26	5006	28.7	94.7	23	5006	34.3	132	28
5007	R_G Range of Pool	24.3 - 56.1											
	D_{max} Range of Pool	71 - 202											
	FOXS Chi^2 Range of Pool	1.5 - 373				2.5 - 1057				3.9 - 2156			
	Chi^2 of best model/ensemble	1.2				1.2				1.4			
	Pool ID/R_G/Dmax/%	5007	30.3	95.7	43	5007	31.4	101	72	5007	32.1	114	75
		5007	28.0	85.9	34	5007	37.8	118	18	5007	27.7	95.3	21
		5007	31.6	103	23	5007	27.8	89.1	9	5007	37.8	123	5

well as more extended conformations (Figure B-4C, E, Table B-2). MES analyses of the extended data subsets in contrast yielded ensembles dominated by more extended structures that do not resemble the crystal structure (Figure B-4E). Taken together, this analysis indicates that the ASPP-PP-1c complexes adopt conformations in solution that can sample the compact conformations observed in the crystal structure. However, these compact conformations interconvert with more extended conformations that may involve release of one or more of the RVxF, SILK or PxxPxR contacts. Very similar results were obtained with ASPP2-PP-1c, suggesting that ASPP2-PP-1c may exhibit similar conformational flexibility in solution (Figure B-5, 6, Table B-3). We next tested the role of the SILK motif in the conformational dynamics of the iASPP-PP-1c complex. We expressed and purified a PP-1c complex containing iASPP(621-828) that lacks the SILK motif and analyzed this complex by SEC-SAXS. Similar to the SILK-containing complexes, this complex eluted as a single peak containing the heterodimer as assessed by MALS (Figure B-7A). Evaluation of the R_G across the peak revealed that this sample is in slow conformational equilibrium between extended states near the peak to more compact states in the tail. Deconvolution of the data and minimal ensemble modeling suggests that the compact forms dominating in the tail of the peak correspond to the crystal structure, while the extended conformations that dominate near the peak maximum may correspond to states in which either the RVxF or PxxPxR motifs have released. Together, these results indicate that while ASPP-PP-1c complexes can sample compact conformations that resemble the iASPP-PP-1c crystal structure, in solution they exhibit conformational flexibility and populate more highly extended forms that likely involve dynamic disengagement of one or more of the discrete contact surfaces.

B.3.4. ASPP proteins modulate PP-1c catalytic activity

To test the ability of ASPP proteins to regulate PP-1c catalytic activity, we first tested the effect of the ASPPs on the activity of PP-1cα towards the chemical PP-1c substrate, para-nitrophenylphosphate (pNPP) (Figure B-8A). pNPP is a small molecule substrate that interacts directly with the PP-1c active site and has been used as a valuable research tool for investigating the effects on the PP-1c active site upon the binding to regulatory proteins and toxins (Choy et al., 2014; Ragusa et al., 2010; Shopik et al., 2013). Both iASPP and ASPP2 were found to increase the activity of PP-1c compared to either a BSA control or the PP-1c inhibitory protein, Inhibitor-

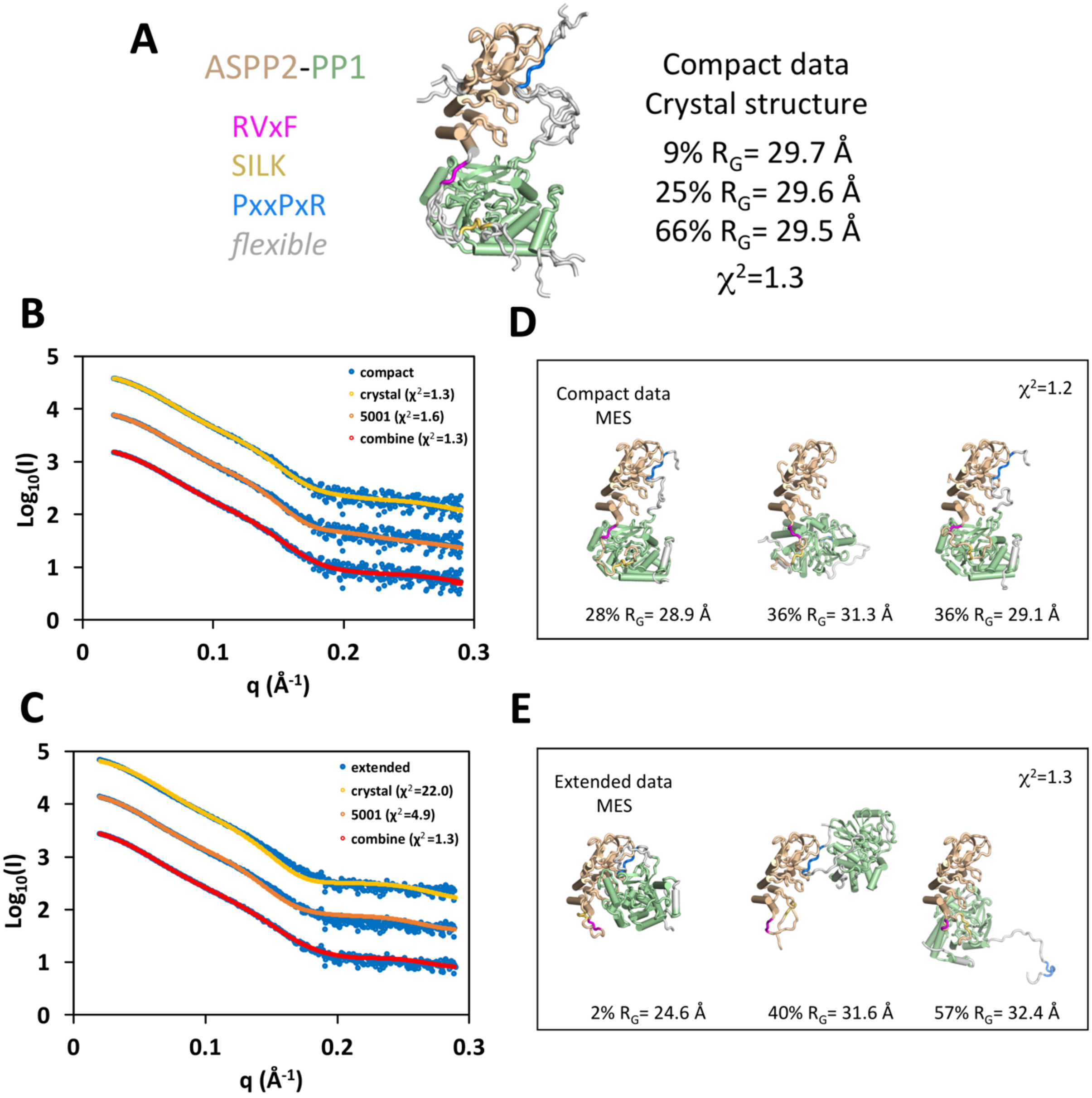


Figure B-6. MES analysis of ASPP2-PP-1c.

Related to Figure B-4. (A) MES modeling using the 5011 library (based on the open form of the iASPP-PP-1c crystal structure). The three model ensemble is shown superimposed with color coding as shown on the left. The percent of each model, R_G and χ^2 are given. (B) and (C) Fit of the compact (B) and extended (C) data sets to calculated scattering from the MES models derived from the 5011 library (crystal), the 5001 library, and the combined libraries. (D) and (E) Three model ensembles derived from the combined libraries for the compact data set (D) and the extended data set (E).

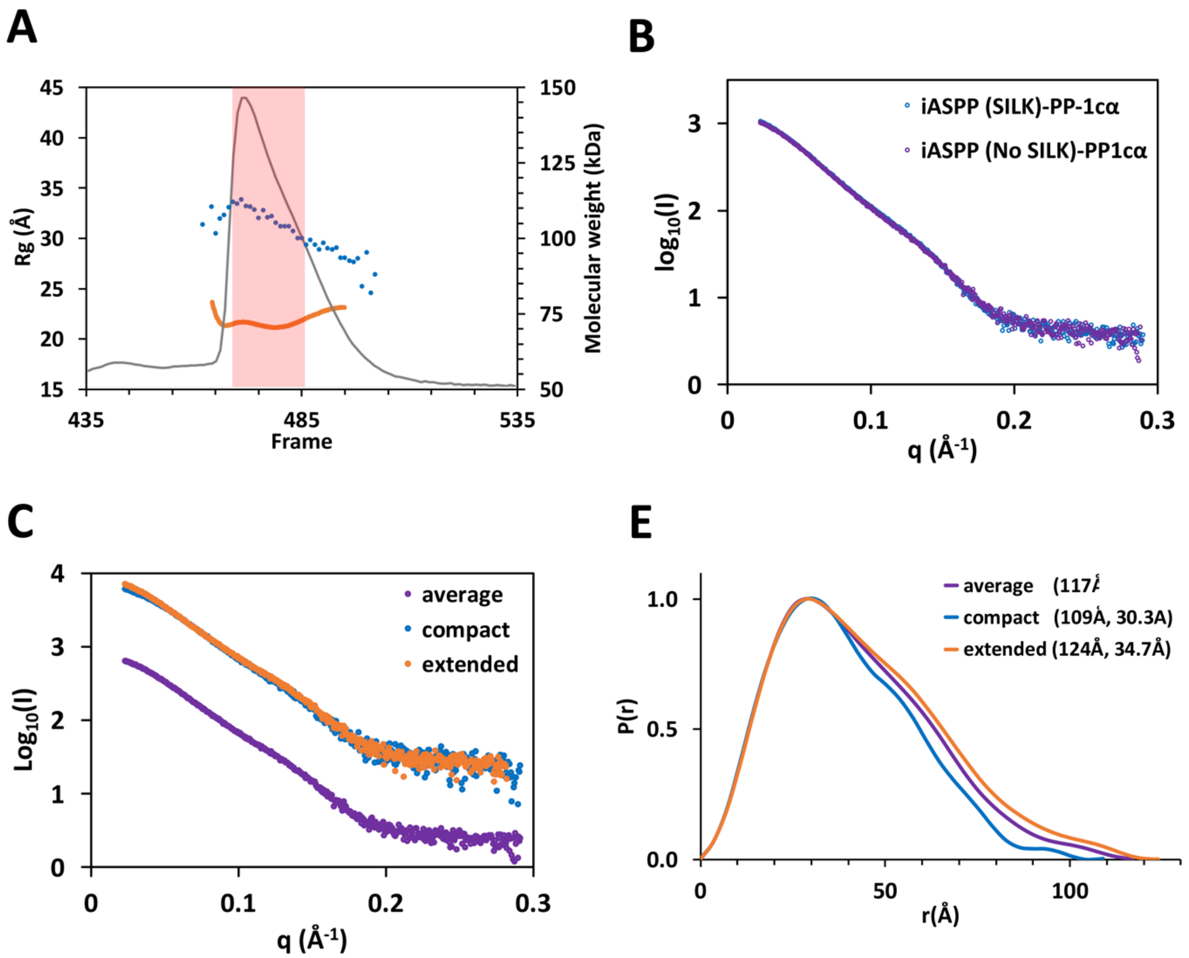


Figure B-7. SEC-SAXS analysis of iASPP(621-828) lacking the SILK motif bound to PP-1c.

Related to Figure B-4. (A) SEC-SAXS of iASPP(621-828)-PP-1c. The grey trace corresponds to the size exclusion UV_{280} trace, the blue dots give the R_G values calculated by Guinier approximation for each frame, the orange trace gives the MW as determined by MALLS. The data from the region marked in pink was used for subsequent analyses. (B) Comparison of scattering from the iASPP(608-828)-PP-1c complex (contains the SILK motif) with scattering from iASPP(621-828)-PP-1c (lacking the SILK motif). (C) RAW deconvolution of iASPP(621-828)-PP-1c scattering into extended and compact subsets. (D) P(r) plots for RAW deconvoluted scattering for iASPP(621-828)-PP-1c. For each trace, the D_{max} (left value) and R_G (right value) are listed.

2 (Figure B-8A). Deletion of the PP-1cα C-terminal tail (PP-1cα 1-300), which is critical for specific interactions with iASPP, effectively blocked the ability of iASPP to enhance PP-1cα dephosphorylation of pNPP, while mutation of the iASPP RVxF motif (L625A) had little effect

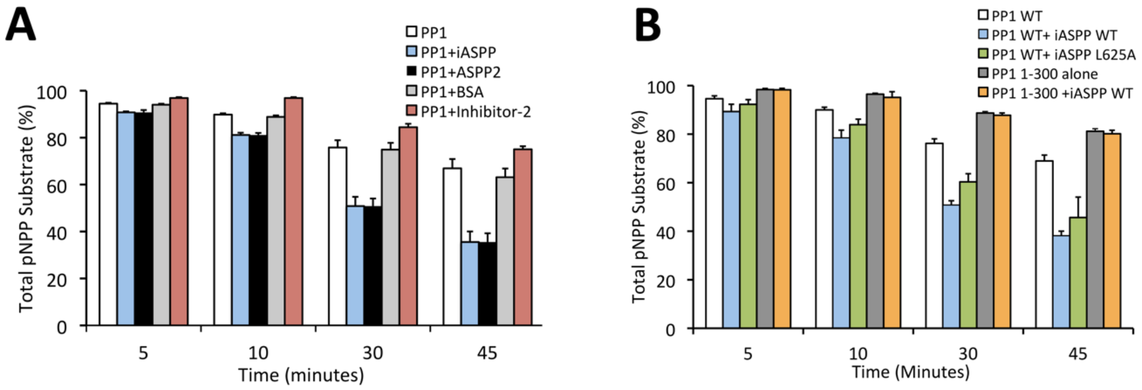


Figure B-8. iASPP and ASPP2 enhance the activity of PP-1c towards pNPP and inhibit PP-1cα towards Phosphorylase a.

(A) PP-1cα colorimetric assay using para-nitrophenylphosphate (pNPP) as a substrate for PP-1cα. PP-1cα was incubated alone or in the presence of iASPP$_{608\text{-}828}$, ASPP2$_{905\text{-}1128}$, bovine serum albumin (BSA) or Inhibitor-2. Graph depicts the total amount of pNPP substrate left in each sample at each time point. Error bars indicate s.d of 3 replicates. (B) PP-1c colorimetric pNPP assay comparing the activity of PP-1cα WT or PP-1cα$_{1\text{-}300}$ alone, or in the presence of WT iASPP$_{608\text{-}828}$, or iASPP$_{608\text{-}828}$ L625A mutant. Error bars indicate s.d of 3 replicates.

on the stimulatory activity of iASPP (Figure B-8B). Taken together, these data indicate that the specific interactions between ASPP proteins and PP-1c enhance the core catalytic activity of PP-1c.

We next addressed the effect of the ASPP proteins on p53 dephosphorylation by PP-1c. We set up an in vitro system using purified recombinant proteins to monitor site-specific dephosphorylation of p53 (Materials and Methods in section B.5.). We chose to monitor the dephosphorylation of p53 on Ser15, a phosphorylation site known to be involved in the regulation of p53 activity and previously shown to be dephosphorylated by PP-1cα both in vitro and in vivo (D. W. C. Li et al., 2006). DNA-dependent protein kinase (DNA-PK), a Ser/Thr kinase that is activated upon binding to double-stranded DNA during the cellular response to DNA damage, was used to phosphorylate p53 on Ser15, as described previously (Lees-Miller et al., 1992; Soubeyrand et al., 2004). Supporting and extending previous results (D. W. C. Li et al., 2006), we found that purified PP-1cα alone could dephosphorylate p53 phospho-Ser15 over a 45 min time course under our reaction conditions (Figure B-9A and B). The addition of an equivalent amount of $iASPP_{608-828}$ or $ASPP2_{905-1128}$ markedly increased the rate of dephosphorylation.

In order to identify the molecular interactions between iASPP and PP-1cα that are important for the ASPP-mediated dephosphorylation of p53, we compared the effect of wild-type and mutant iASPP and PP-1cα proteins on the dephosphorylation of p53 Ser15 (Figure B-9C-H). To test the importance of the iASPP RVxF-like motif ($_{622}RARL_{625}$), we tested the effect of the iASPP L625A mutant on p53 dephosphorylation. This mutation resulted in a moderate reduction p53 Ser15 dephosphorylation, compared to WT iASPP (Figure B-9C and D). We also tested the importance of the iASPP ANK-PP-1c pivot interaction by creating an iASPP D663R mutant, which reverses the charge of a key residue forming an electrostatic contact with residue Arg261 of PP-1cα (Figure B-3E). This mutation exhibited a similar defect in p53 dephosphorylation as iASPP L625A (Figure B-9C and D). We next tested the effects of two mutations in the PP-1cα C-terminal tail that is essential for binding to the SH3 domain of ASPP proteins: PP-1cα1-300, which lacks the entire C-terminal tail, and PP-1cα T320D, a mutant that mimics phosphorylation on Thr320 that has been associated with PP-1c autoinhibition (Berndt et al., 1997; Dohadwala et al., 1994; Skene-Arnold et al., 2013). Despite the fact that these mutations have been shown to destabilize iASPP-PP-1c interactions (Skene-Arnold et al., 2013), mutation of PP-1cα T320D had no apparent effect on iASPP-mediated dephosphorylation of p53 and the PP-1cα1-300 mutant only marginally decreased the dephosphorylation of p53 Ser-15 (Figure B-9E and F). Combining either of the PP-1cα tail mutations with either of the iASPP mutations further decreased but did not abolish the iASPP effect (Figure B-9C-H). Taken together, these data indicate that ASPP proteins can facilitate dephosphorylation of a critical phosphosite in p53 in a manner that relies on the integrity of multiple contacts between ASPP and PP-1cα.

B.3.5. Differential inhibition of p53 DNA binding by ASPP and ASPP-PP-1c complexes

ASPP2 has long been known to bind the p53 DNA binding domain (DBD) through interactions between the ASPP2 SH3 domain and the DNA binding loops of p53 (Gorina & Pavletich, 1996; Patel et al., 2008). The ability of the ASPP proteins to specifically bind p53 suggests that the ASPP proteins could act as a targeting protein, bringing p53 to the vicinity of PP-

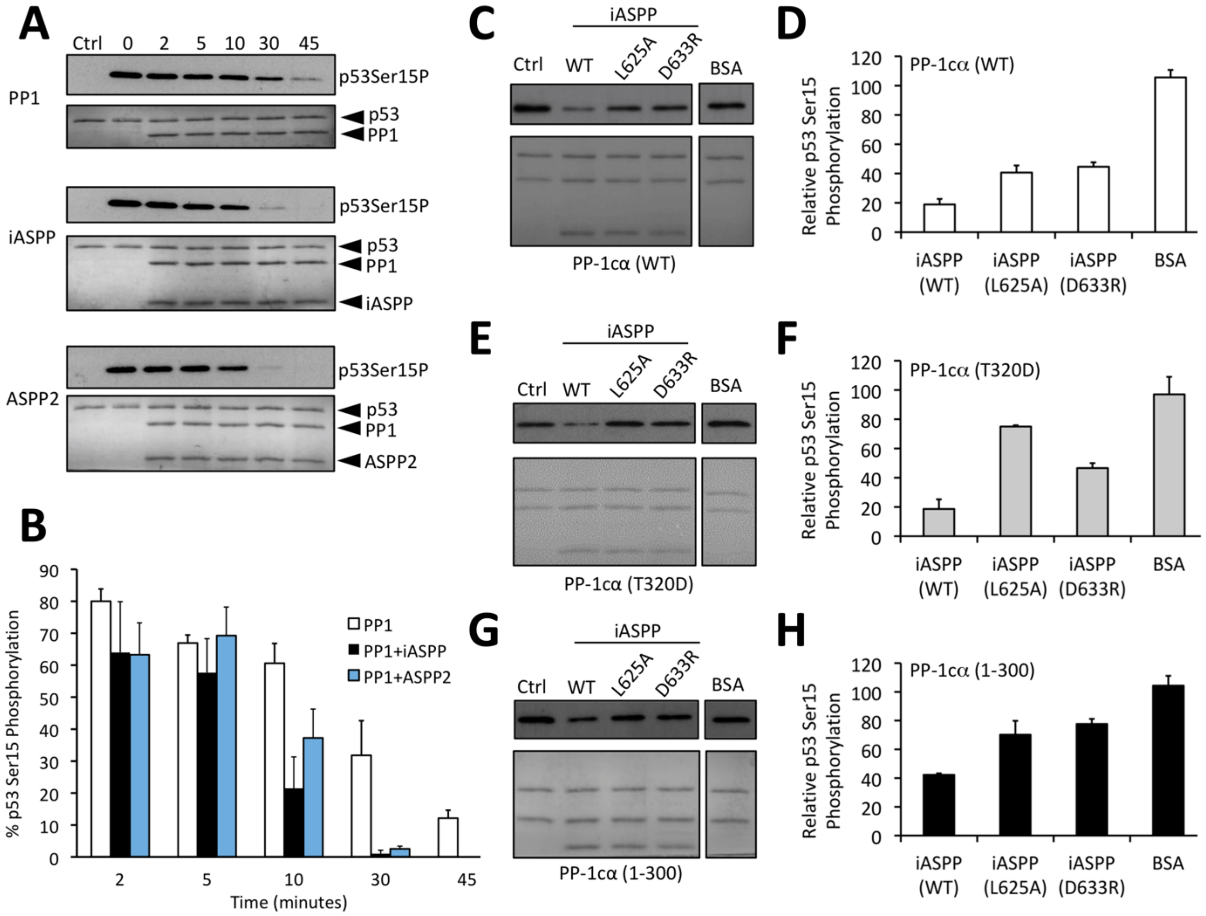


Figure B-9. iASPP and ASPP2 enhance the dephosphorylation of p53 Ser-15.

(A) Western blot analysis of in vitro dephosphorylation reactions of DNA-PK phosphorylated p53$_{2-293}$ (blotting for phosho-Ser-15) incubated with PP-1cα alone (top panel) or in the presence of iASPP$_{608-828}$ (middle panel) or ASPP2$_{905-1128}$ (lower panel) for 0 to 45 min. The lanes labeled "ctrl" contain a sample of p53 where DNA-PK was inhibited with LY294002 for the duration of the experiment. The lanes labeled "0" contain no PP1 or ASPP-PP1. (B) Quantification of the amount of phosphorylated p53 Ser-15 in each sample relative to the total phospho-Ser-15 in control. Error bars indicate s.d. of 4 replicates. 1(C-H) Western blot analyses of in vitro dephosphorylation reactions of DNA-PK phosphorylated p53 (blotting for phospho-Ser-15 p53) by PP-1cα WT (C and D), PP-1cα T320D (E and F), or PP-1cα$_{1-300}$ (G and H) in the presence of iASPP WT, iASPP L625A, or iASPP D633R, with results quantitated as in (B). Error bars indicate s.d. of 3 replicates.

1c for dephosphorylation (Figure B-10A). While the striking similarity of iASPP to ASPP2 suggests that similar interactions may occur between iASPP and p53 DBD, other studies have indicated that the poly-proline regions in p53 may also bind the iASPP SH3 domain (Ahn et al., 2009; Bergamaschi et al., 2006). Alignment of iASPP-PP-1c on the ASPP2-p53 structure reveals that while the overall p53 binding surface is similar, limited substitutions in the binding interface might lead to different binding properties of the two ASPP proteins (Figure B-10B). The fact that the p53 DBD can potentially bind either DNA or the ASPP SH3 presents the possibility that these interactions could be mutually exclusive.

To directly test this idea, we used an electrophoretic mobility shift assay to examine the effects of either free ASPP or ASPP-PP-1c on the binding of the p53 DBD to a consensus p53 DNA recognition sequence (Figure B-10C). The results show that the p53-DNA complex is largely resistant to challenge by iASPP, and competition is only observed at the highest iASPP concentration. In contrast, ASPP2 competes p53 off the DNA at much lower concentrations, suggesting that ASPP2 SH3 binds more tightly to the p53 DBD than iASPP. Complex formation between either ASPP protein and PP-1c abrogates the p53 inhibitory effect of the ASPP proteins, consistent with the idea that the ASPP SH3 peptide binding groove is critical for both p53 and PP-1c binding.

B.4. Discussion

Here we present a structural model for interactions between iASPP and PP-1c. Our model suggests that interactions between these proteins are stabilized by multiple discrete contacts that together constitute a large binding interface that can explain the high affinity of this complex (KD ~ 26 nM) (Skene-Arnold et al., 2013). These include the docking of the PxxPxR motif of the PP-1c tail onto the ASPP SH3 domain, docking of iASPP SILK-like and RVxF-like motifs onto the PP-1c catalytic domain, as well as docking of the N-terminal iASPP ANK motif against the PP-1c catalytic domain. The high degree of sequence conservation within the ASPP family suggest that ASPP1/2 may bind PP-1c in a very similar manner, and indeed SAXS analysis indicates that the structure and dynamics of the ASPP2-PP-1c complex is similar to iASPP-PP-1c in solution (Figure B-5, 6). While this manuscript was under review, Bertran et al. published a paper reporting on the crystal structure of the ASPP2-PP-1c complex (Bertran et al., 2019). The structure is very similar to the open form of iASPP-PP-1c, except that the ASPP2 construct used by Bertran lacks the SILK motif. These authors went on to test the binding affinities of the three ASPP variants with the three

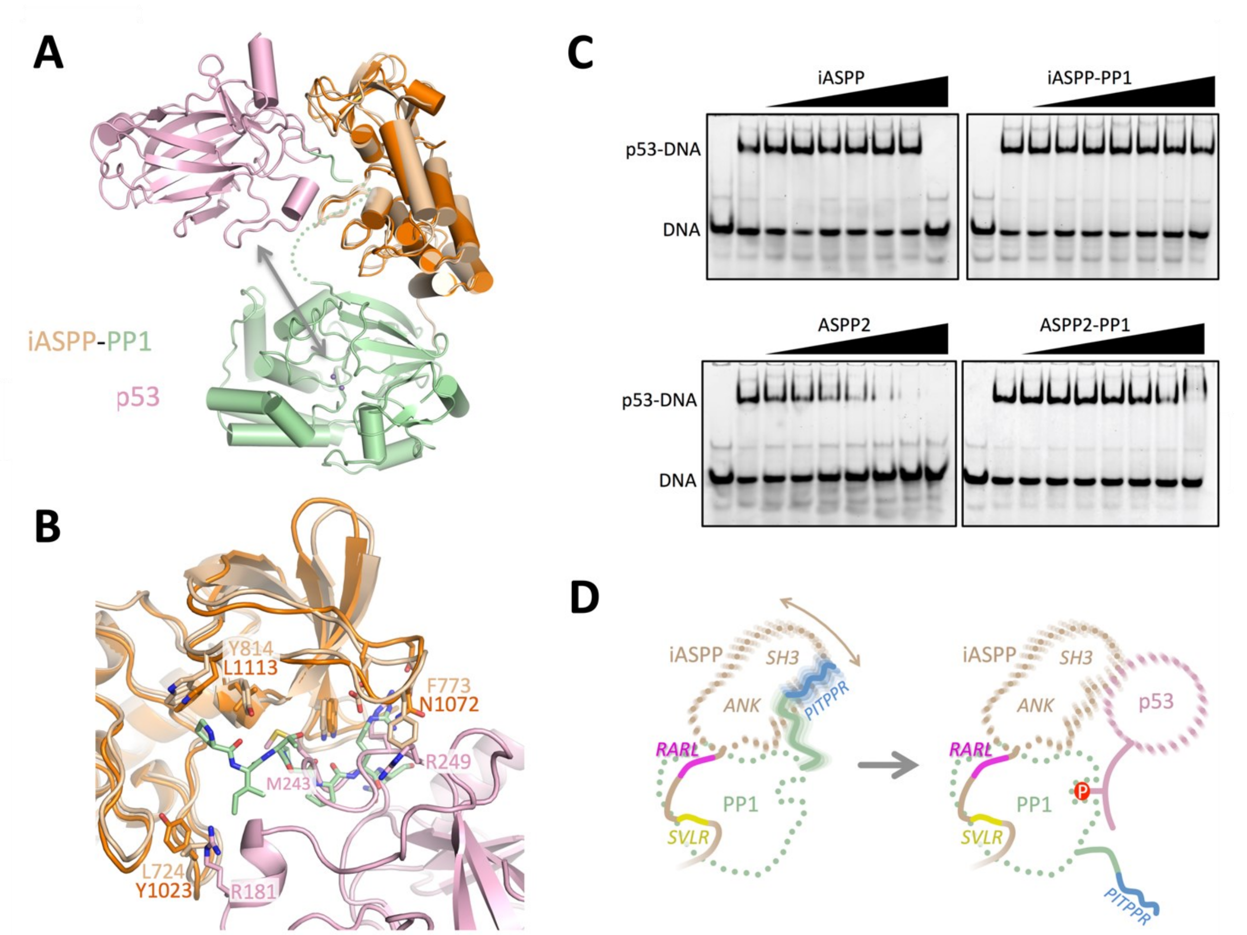


Figure B-10. ASPP-PP-1c-p53 interactions and a model for the targeted dephosphorylation of p53 by ASPP-PP-1c.

(A) Interactions between ASPP-PP-1cα complexes and p53. Shown is a structural overlay of the closed iASPP-PP-1cα complex and the ASPP2-p53 DBD complex (PDB ID: 1YCS), aligned on their ASPP components. The arrow indicates the proximity of the p53 DBD to the PP-1c active site catalytic metal ions (purple spheres). (B) Details of the superposition of the iASPP-PP-1c and ASPP2-p53 DBD complex. Shown is the ASPP SH3 domain and C-terminal ANK finger with the p53 DBD and PP-1c PxxPxR peptide shown. Key residues involved in intermolecular contacts are shown as sticks. (C) EMSA analysis probing the ability of ASPP or ASPP- PP-1cα complexes to inhibit p53 DBD binding to a PUMA p53 DNA site. Each lane contains 80 nM PUMA. In each panel, lanes 2-9 contain 800 nM p53 DBD, and a gradient of ASPP protein or ASPP- PP-1cα complex at 0, 0.8, 1.6, 3.2, 6.4, 12.8, 25.6, and 51.2 μM. (D) Model for how a dynamic ASPP-PP-1c complex could bind and target p53 for dephosphorylation. Left panel – the iASPP-PP-1c complex is stabilized by discrete interactions involving the SVLR, RARL and PITPPR peptide motifs, resulting in a flexible complex that allows motion of the iASPP ANK/SH3 domain relative to the PP-1c catalytic domain. Right panel – dynamic release of the PP-1c PITPPR motif allows binding of p53 to the ANK/SH3 domain and dephosphorylation.

PP-1c variants. They discovered that while all pair-wise complexes are possible, the binding affinities of different pairs can vary by almost 100-fold. ASPP2 showed a significant preference for PP-1cα, and this selectivity could be attributed to interactions between charged residues immediately C-terminal to the PP-1cα PxxPxR motif and the ASPP SH3 domain (Bertran et al., 2019).

The interactions of the RVxF and ANK repeats of iASPP with PP-1c are reminiscent of the interactions with PP-1c of the myosin binding protein and PP-1cα regulatory subunit, MYPT1 (Terrak et al., 2004) (Figure B-1D). The orientation of the N-terminal ANK repeats of MYPT1 are similar to the more closed conformation of the iASPP-PP-1c complex in which the fingers of the two N-terminal-most ANK repeats contact the PP-1c catalytic domain. MYPT1, however, does not contain an SH3 domain but contains additional C-terminal ANK repeats that wrap around the PP-1c C-terminal tail, potentially explaining the specificity of MYPT1 for PP-1cβ (Terrak et al., 2004)(Figure B-1D).

Intriguingly, analysis of either iASPP-PP-1c or ASPP2-PP-1c by SEC-SAXS indicates a surprising degree of conformational heterogeneity in both of these complexes (Figures B-4, 6, 11). Our analysis and structural modeling based on these data suggests that the conformation represented by the iASPP-PP-1c crystal structure is in slow exchange with more extended conformations under physiological solution conditions. Notably, the more extended conformations cannot be fit by simply allowing rigid body movement of the PP-1c catalytic domain relative to the iASPP ANK/SH3 domain (Figure B-4D). However, our analysis suggests that these more extended states could be accessed through the release of one or more of the contacts involving the discrete peptide-domain interactions such as the PxxPxR-SH3 interaction, or the interactions between the SILK or RVxF motifs with the PP-1c catalytic domain. Each of these individual peptide-domain interactions are expected to be of relatively low affinity compared to the overall high affinity of the complex, and are therefore likely to be microscopically dynamic. Indeed, SAXS analysis has indicated that a similar dynamic release of SILK and RVxF motifs are involved in the formation of a heterotrimeric complex between spinophilin, inhibitor 2 and PP-1c (Dancheck et al., 2011). Despite the fact that both spinophilin and inhibitor 2 have been shown to bind to PP-1c primarily via a RVxF motif (Hurley et al., 2007; Ragusa et al., 2010), within the heterodimeric complex, only the RVxF motif from spinophilin was bound to PP-1c.

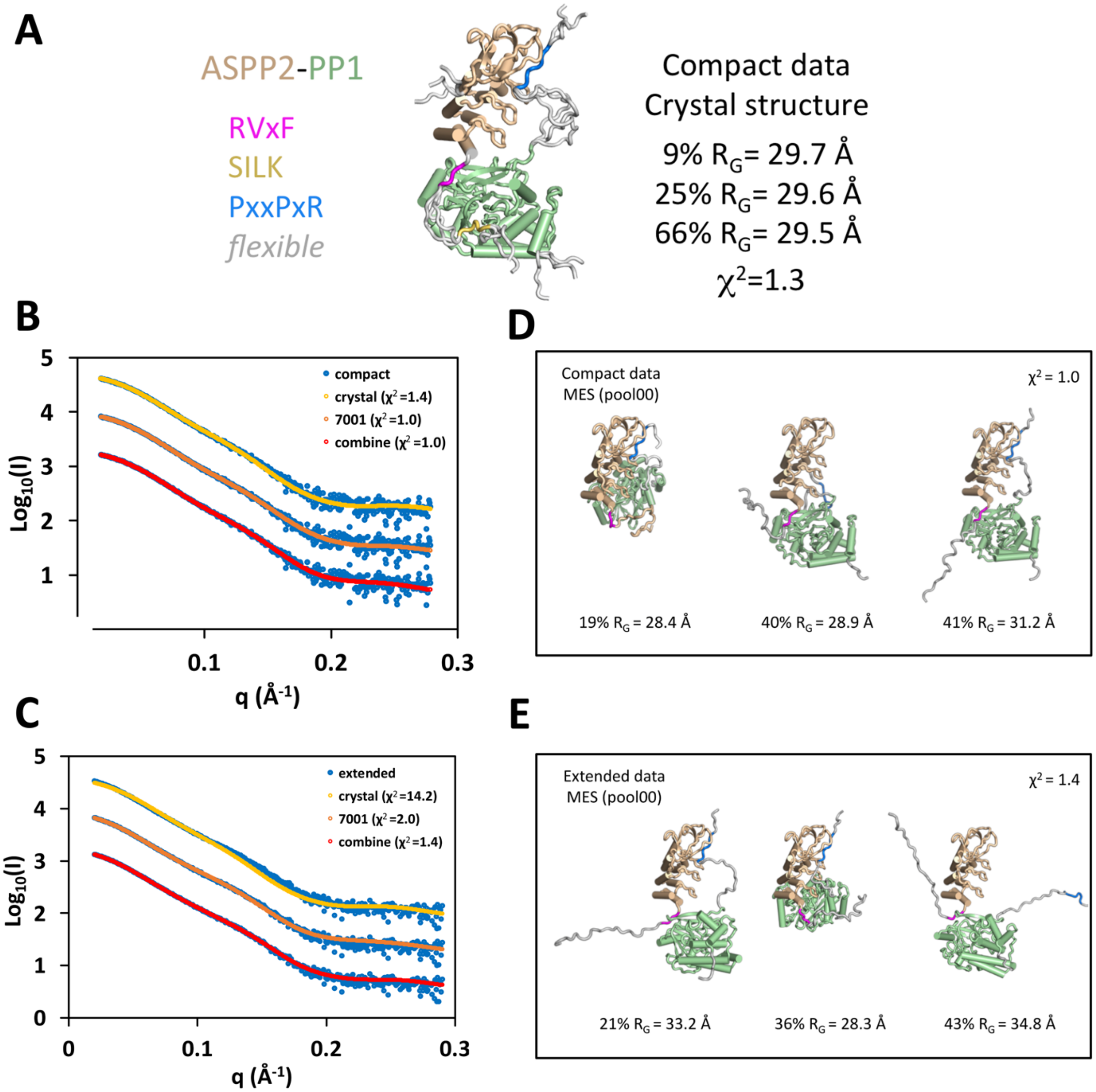


Figure B-11. MES analysis of iASPP(621-828) lacking the SILK motif bound to PP-1c.

Related to Figure B-5. (A) MES modeling against the compact iASPP(621-828)-PP-1c data set using a library based on the open form of the iASPP-PP-1c crystal structure with the SILK motif deleted. The single best fit model is shown with color coding as shown on the left. (B) and (C) Fit of the compact (B) and extended (C) data sets to calculated scattering from the MES models derived from the crystal structure library, the 7001 library (in which the ASPP and PP-1c domains are defined as independent rigid bodies), and the combined libraries, in which the RVxF and PxxPxR motifs are allowed to release from their respective binding partners. (D) and (E) Three model ensembles derived from the combined libraries for the compact data set (D) and the extended data set (E). The percent of each model, R_G and χ^2 are given.

The work here also has implications for the functional role of the ASPP proteins in the regulation of PP-1c activity. We found that ASPP proteins can activate PP-1c catalytic activity against the small molecule substrate pNPP as well as the physiologically relevant Ser15 phosphosite in p53 (Figure B-8, 9). The ASPP proteins, like the MYPT proteins have been implicated as PP-1c targeting proteins that might act to recruit specific substrates to PP-1c for dephosphorylation (Helps et al., 1995; Llanos et al., 2011; Skene-Arnold et al., 2013). ASPP proteins are not only known to specifically bind to p53 using their ANK-SH3 domain (Helps et al., 1995; Llanos et al., 2011; Skene-Arnold et al., 2013), but they also can bind to a variety of other potential targets. These include NF-kB p65 (Yang, Hori, Sanda, et al., 1999; Yang, Hori, Takahashi, et al., 1999), Yes-associated protein YAP (Espanel & Sudol, 2001), and Bcl-2 (Katz et al., 2008; Naumovski & Cleary, 1996), many of which may also be regulated by PP-1c (Mui et al., 2015; Vigneron et al., 2010). While there is no high resolution structural information for these other interactions, it is likely that the SH3 peptide binding groove is involved in at least some of these complexes.

Together with our previous binding studies (Skene-Arnold et al., 2013), our crystal structure demonstrates that the PxxPxR motif in the PP-1c tail is critical for ASPP-PP-1c complex formation. This interaction however occludes the SH3 surface that is critical for binding to the p53 DNA binding domain, and indeed we show that the binding of PP-1c to ASPP reduces the ability of ASPP proteins to inhibit p53 DNA binding (Figure B-10). This finding then raises the question as to how ASPP proteins could play a targeting role if their primary target binding surface, the SH3 domain, is occluded by PP-1c. Our combined results support the possible dynamic release of the SH3 domain by the PxxPxR peptide, as suggested by our SAXS analysis. This dynamic interface plasticity might free up the SH3 domain to bind protein targets for dephosphorylation by PP-1c (Figure B-10D). The additional flexibility afforded by the transient dissociation of the PxxPxR-SH3 interaction may furthermore facilitate dephosphorylation of diverse phosphosites.

B.5. Materials and methods

Purification of ASPP-PP-1c for crystallization and small angle X-ray scattering.

iASPP$_{608\text{-}828}$ bacterial pellets from 4 L culture were resuspended in 100 mL lysis Buffer A (25 mM Tris, pH 8.0, 500 mM NaCl, 10% glycerol) containing protease inhibitor cocktail (Pierce Halt). Cells were lysed with three passes through an Emulsiflex-C3 high pressure homogenizer (Avestin). Lysates were clarified by centrifugation at 13000g for 30 minutes at 4 °C. The iASPP supernatant was loaded onto a 4 mL pre-equilibrated Ni-NTA sepharose column, followed by washing with 150 mL wash buffer 1 (25 mM Tris, pH 8.0, 500 mM NaCl, 10 mM imidazole). iASPP-bound Ni-NTA beads were incubated with PP-1cα at 4 °C for 1 hour. The supernatant was removed, then the beads were washed with 150 mL wash buffer 1 and 100 mL wash buffer 2 (25 mM Tris, pH 8.0, 500 mM NaCl, 20 mM imidazole). The iASPP His-tag was removed by on-column digestion with 300 μL 6 mg/mL GST-3C protease in digestion buffer (25 mM Tris, pH 8.0, 150 mM NaCl, 2 mM DTT) at 4°C overnight. The cleaved iASPP-PP-1cα complex was further purified by Q sepharose chromatography with Buffer A (40 mM Tris, pH 8.0, 20 mM NaCl, 1 mM DTT) and Buffer B (40 mM Tris, pH 8.0, 1 M NaCl, 1 mM DTT). iASPP-PP1cα was finally purified by size exclusion chromatography using a Superdex 75 16/60 column with Buffer S (25 mM Tris, pH 8.0, 150 mM NaCl, 1 mM DTT). The protein complex was exchanged into crystallization Buffer C (25 mM Tris, pH 8.0, 100 mM NaCl, 1 mM TCEP, 5% glycerol), concentrated to 4 mg/mL. iASPP621-828 and ASPP2905-1128 were purified similarly.

Crystallization and structure determination of iASPP-PP-1cα.

Crystals of iASPP608-828-PP-1cα complex were grown by sitting drop vapor diffusion with a reservoir solution containing 12-15% PEG6000, 0.1 M sodium citrate, pH 5.3 and were improved by seeding by hanging vapor diffusion at room temperature.

For data crystallographic data collection, iASPP$_{608\text{-}828}$-PP-1cα crystals were flash frozen in liquid nitrogen with a cryoprotectant (reservoir solution with 30% glycerol). Diffraction data were collected on beamline 12.3.1 at the Advanced Light Source and indexed, integrated and scaled using HKL2000. Two data sets were collected, exposing different parts of a single crystal.

An automated combinatorial algorithm was used to select the range of frames to maximize data redundancy while excluding frames suffering from radiation damage. Molecular replacement was carried out using PHASER, placing two copies of human PP-1 catalytic subunit (search model PDB ID 3EGG, chain A) and two copies of human iASPP$_{607\text{-}822}$ (PDB ID 2VGE). Refinement in PHENIX utilized automated non-crystallographic symmetry and reference model restraints throughout with the latter being removed for the final refinement round. The PxxPPR, RARL and SILK motif regions were sequentially built toward the end of refinement when the difference density had improved sufficiently to be confident of placement. Careful analysis of binding modes in homologous PxxPPR, RARL and SILK motif complex structures were used to guide sequence register and orientation.

Size exclusion chromatography – small angle X-ray scattering (SEC-SAXS) analysis of ASPP-PP-1c complexes.

SEC-SAXS and multi-angle laser light scattering (SEC-MALLS) was performed on iASPP$_{608\text{-}828}$-PP-1c, iASPP$_{621\text{-}828}$ and ASPP2$_{905\text{-}1128}$–PP-1c complexes at beamline 12.3.1 at the Advanced Light Source (ALS). iASPP-PP-1c and ASPP2-PP-1c samples at either 10 mg/mL and 15 mg/mL were loaded onto a Shodex KW-803 size exclusion column (Shodex) in 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM DTT buffer running at 0.5 mL/min. The column eluant passed through a splitter and was simultaneously analyzed for SAXS on the beamline using a PILATUS3 X 2M detector and MALS using a DAWN detector (Wyatt Technology). MALS data was analyzed using ASTRA VI software (Wyatt Technology) to provide MW information across the eluted peak. SAXS data was analyzed with RAW (Hopkins et al., 2017) to derive RG values for individual 3 sec. frames across the peak using the Guinier approximation.

RAW was used for deconvolution of iASPP-PP1c and ASPP2-PP1c data into separate compact and extended subsets. The range of data selected for the deconvolution extended from the peak to ½ maximal peak height in the tail, a region where MALLS indicated the sample only contained ASPP-PP-1c heterodimer and not higher order aggregates. Each of these datasets was further processed using SCATTER (https://bl1231.als.lbl.gov/scatter/) to determine R_G by Guinier analysis for data truncated to q x RG < 1.3 and D_{max} by P(r)-distribution plot (Rambo & Tainer, 2013).

SAXS-based modeling of ASPP-PP-1c structure and dynamics using a minimal ensemble search (MES).

BILBOMD was used to generate model libraries for MES analysis (Pelikan et al., 2009). The starting PDB file for BILBOMD was based on the crystal structure with the flexible regions of peptide not visualized in the crystal structure modeled using SWISS-MODEL (https://swissmodel.expasy.org) and AllosMod-FoXS (https://modbase.compbio.ucsf.edu/allosmod-foxs/). Several model sets with different levels of flexibility were created. The first library (the 5011 library) was designed to represent the open crystal form. In this library, the iASPP ANK-SH3 domain, the PP-1c catalytic domain, as well as the RVxF, SILK and PxxPxR peptides of the open crystal form were treated as a single rigid body, while the peptide linkers and tails that were not observed in the crystallographic electron density were defined as flexible. Library 5013 is analogous to 5011, and uses instead the closed conformation observed crystallographically. Library 5001 allowed additional flexibility by treating the iASPP ANK-SH3 domain and the PP-1c catalytic domain as independent rigid bodies. While these libraries could model the compact datasets generated by RAW, the extended data sets were not well fit by models from these libraries. To allow further flexibility to model the extended ASPP-PP-1c structures observed in solution, we created several additional model libraries in which one or more of the RVxF, SILK or PxxPxR peptides were released from their respective binding partners for $iASPP_{608-828}$ and $ASPP_{905-1128}$. These model libraries provided much larger degrees of flexibility as expressed by the range of RG and Dmax for the models in these libraries. In the 5002 library, the PxxPxR peptide was released from its interaction with the iASPP SH3, while the RVxF and SILK interactions were maintained. In the 5003 library, the SILK interaction was released, while the RVxF and PxxPxR contacts were maintained. In the 5004 library, the RVxF interaction was released, while the SILK and PxxPxR contacts were maintained. In the final libraries, two of the peptide contacts were released, leaving only a single peptide interaction to tether the iASPP and PP-1c domains. In library 5005, the two proteins were tethered by the SILK interaction, in library 5006, the tether was provided by the RVxF, while in library 5007 the PxxPxR provided the tethering interaction. We also generated library for iASPP$_{621\text{-}828}$-PP-1c which lacks SILK motif on iASPP. Library 7001 for iASPP$_{621\text{-}828}$-PP-1c were created with restraints corresponding to library 5001. In the library 7002, the PxxPxR peptide was released from its interaction with the iASPP SH3, while the RVxF interactions were fixed. In the 7003 library, the RVxF contacts with the PP-1c were released, while the PxxPxR peptide interaction was maintained. MES was performed on each of the individual model sets, as well as on all of the model sets combined as previously described (Daniel Aceytuno et al., 2017; Hodge et al., 2016; Pelikan et al., 2009). Details of the model libraries and the MES results for each of the model pools are given in Table B- for iASPP$_{608\text{-}828}$-PP-1c and Supplementary Table 2 for ASPP2$_{905\text{-}1128}$-PP-1c.

pNPP Phosphatase Activity Assay (Zhuo et al., 1993)**.**

Colorimetric phosphatase activity assays using para-nitrophenylphosphate (pNPP) were carried out as follows. ASPP2$_{905\text{-}1128}$, WT iASPP$_{608\text{-}828}$, iASPP$_{608\text{-}828}$ L625A mutant, Bovine Serum Albumin, and inhibitor-2 proteins were diluted in Buffer L (50 mM Tris-HCl, pH 8.0, 75 mM KCl, 10 mM NaCl, 10 mM $MgCl_2$, 0.2 mM EGTA, 1 mM $MnCl_2$ and 0.2 % β-mercaptoethanol) to a final concentration of 500 nM. PP-1cα WT and PP-1cα 1-300 was diluted in buffer L to 333 nM and added together with each regulatory protein in a 96 well plate. The samples were then incubated for 10 min and reaction was initiated by addition of pNPP (1.2 mM final concentration). Reaction was monitored by absorbance at 405 nm over a 45 min. time course at 37 °C. The percentage of total pNPP substrate was calculated from the raw absorbance of each reaction at a given time point (Abx) and the total absorbance after ~5 hrs (Ab_{total}) as follows: $[(Ab_{total} - Ab_x) / Ab_{total}]$. Each experiment was carried out three times in duplicate and the standard deviation is shown as error bars.

***In Vitro* p53 Dephosphorylation Assay by Protein Phosphatase-1cα.**

Purified p53$_{2\text{-}293}$ (0.5 μg, 940 nM final concentration) was phosphorylated with 75 U of DNA-PK (Promega) in DNA-PK kinase buffer (50 mM Tris-HCl, pH 8.0, 75 mM KCl, 10 mM NaCl, 10 mM $MgCl_2$, 1 $MnCl_2$ 0.2 mM EGTA, 2 mM DTT, 0.5 mM ATP, 10 μg/mL sonicated calf thymus DNA, 0.025% Tween-20) for 90 minutes at 30 °C and stopped with 1 mM LY294002 (Calbiochem). PP-1cα, either alone or combined with ASPP, was pre-incubated at 30 °C for 15 minutes in DNA-PK kinase buffer and dephosphorylation was initiated by addition of phosphorylated p53 in a 10 µL reaction volume (final concentrations of 470 nM $p53_{2\text{-}293}$, 330 nM PP-1cα, 330 nM PP-1cα (T320D), 360 nM PP-1cα1-300, 520 nM $iASPP_{608\text{-}828}$, or 490 nM $ASPP2_{905\text{-}1128}$). The dephosphorylation reaction was stopped using 2X SDS-PAGE sample buffer.

Dephosphorylation was detected and quantitated by western blotting.

p53 samples were separated by 12% SDS-PAGE and transferred onto nitrocellulose using a Mini-Blot Module (Biorad). Nitrocellulose blots were first stained with Memcode Reversible Protein Stain kit (Pierce Scientific) and then scanned to analyze total amount of protein in each dephosphorylation reaction. After removing the reversible stain, blots were blocked using 5 % milk in TBST buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, and 0.1 % Tween-20) for 1 hr at room temperature. Membranes were then incubated at 4 °C overnight (or for 1 hour at room temperature) with antibody specific for p53 phosphorylated on Ser-15 (Cell Signaling), diluted to 1:15000, in 5 % milk in TBST buffer. Membranes were washed with TBST and further incubated with a horseradish peroxidase-conjugated immunoglobin G secondary antibody (Cell Signaling) for 1 hr at room temperature. Blots were developed for 1 min in enhanced chemiluminescence solution (ClarityTM Western ECL substrate, Biorad) and exposed to Fuji Medical X-Ray Film (Fujifilm). Densitometry for total p53 (stained with Memcode Reversible stain) and phospho-p53 (western blot) were quantified using Image Studio Light or ImageJ software. The amount of phospho-p53 (either p53 phospho-Ser-15) was normalized to the total p53 protein present in each sample. Data was evaluated for outliers by calculating the interquartile range (IQR) and data points were deemed an outlier if they were greater than 1.5 times the IQR. Data was subjected to a one-way analysis of variance (ANOVA) and a post hoc Tukey's test for individual multiple comparison ($p < 0.05$).

Electrophoretic mobility shift assay.

Complementary nucleotides, PUMA forward and reverse strands, were annealed in annealing buffer (10 mM Tris pH 7.5, 100 mM NaCl, 1 mM EDTA) at 95°C for 2 min and slowly cooled to 23oC. Double stranded DNA (final concentration 80 nM) was incubated with p53 DBD (residues 94-292, final concentration 800 nM) on ice for 30 minutes. ASPP or ASPP-PP1c complex were subsequently added at the concentrations indicated (Figure 6) and incubated on ice for 30 minutes. Binding reactions were performed in 40 mM Tris pH 7.5, 150 mM NaCl, 3% glycerol. Samples were separated by electrophoresis on 6% (wt/vol) native polyacrylamide gels in TBE at 4°C. Gels were visualized by a Typhoon scanner (GE HealthCare).

SAXS analyses. Goodness of fit (χ^2) of individual models to experimental scattering data was calculated using FoXS and these values are reported in Figures B-2, 4, 5, 6 and Tables B-2, 3.

pNPP Phosphatase Activity Assay. The percentage of total pNPP substrate was calculated from the raw absorbance of each reaction at a given time point (Abx) and the total absorbance after ~5 hrs (Abtotal) as follows: [(Abtotal – Abx) / Abtotal]. Each experiment was carried out three times in duplicate and the standard deviation is shown as error bars. Results are reported in Figure B-8.

In Vitro p53 Dephosphorylation Assay by Protein Phosphatase-1cα. Densitometry for total p53 (stained with Memcode Reversible stain) and phospho-p53 (western blot) were quantified using Image Studio Light or ImageJ software. The amount of phospho-p53 (either p53 phospho-Ser-15) was normalized to the total p53 protein present in each sample. Data was evaluated for outliers by calculating the interquartile range (IQR) and data points were deemed an outlier if they were greater than 1.5 times the IQR. Data was subjected to a one-way analysis of variance (ANOVA) and a post hoc Tukey's test for individual multiple comparison ($p < 0.05$). Results are reported in Figure 5.

B.6. References

Agirre, X., Román-Gómez, J., Jiménez-Velasco, A., Garate, L., Montiel-Duarte, C., Navarro, G., Vázquez, I., Zalacain, M., Calasanz, M. J., Heiniger, A., Torres, A., Minna, J. D., & Prósper, F. (2006). ASPP1, a common activator of TP53, is inactivated by aberrant methylation of its promoter in acute lymphoblastic leukemia. *Oncogene*, *25*(13), 1862–1870. https://doi.org/10.1038/sj.onc.1209236

Ahn, J., Byeon, I. J. L., Byeon, C. H., & Gronenborn, A. M. (2009). Insight into the structural basis of pro- and antiapoptotic p53 modulation by ASPP proteins. *Journal of Biological Chemistry*, *284*(20), 13812–13822. https://doi.org/10.1074/jbc.M808821200

Ao, Y., Rohde, L. H., & Naumovski, L. (2001). p53-interacting protein 53BP2 inhibits clonogenic survival and sensitizes cells to doxorubicin but not paclitaxel-induced apoptosis. *Oncogene*, *20*(21), 2720–2725. https://doi.org/10.1038/sj.onc.1204352

Aylon, Y., Ofir-Rosenfeld, Y., Yabuta, N., Lapi, E., Nojima, H., Lu, X., & Oren, M. (2010). The Lats2 tumor suppressor augments p53-mediated apoptosis by promoting the nuclear proapoptotic function of ASPP1. *Genes and Development*, *24*(21), 2420–2429. https://doi.org/10.1101/gad.1954410

Bergamaschi, D., Samuels, Y., O'Neil, N. J., Trigiante, G., Crook, T., Hsieh, J. K., O'Connor, D. J., Zhong, S., Campargue, I., Tomlinson, M. L., Kuwabara, P. E., & Lu, X. (2003). iASPP oncoprotein is a key inhibitor of p53 conserved from worm to human. *Nature Genetics*, *33*(2), 162–167. https://doi.org/10.1038/ng1070

Bergamaschi, D., Samuels, Y., Sullivan, A., Zvelebil, M., Breyssens, H., Bisso, A., Del Sal, G., Syed, N., Smith, P., Gasco, M., Crook, T., & Lu, X. (2006). iASPP preferentially binds p53 proline-rich region and modulates apoptotic function of codon 72-polymorphic p53. *Nature Genetics*, *38*(10), 1133–1141. https://doi.org/10.1038/ng1879

Berndt, N., Dohadwala, M., & Liu, C. W. Y. (1997). Constitutively active protein phosphatase 1α causes Rb-dependent G1 arrest in human cancer cells. *Current Biology*, *7*(6), 375–386. https://doi.org/10.1016/s0960-9822(06)00185-0

Bertran, M. T., Mouilleron, S., Zhou, Y., Bajaj, R., Uliana, F., Kumar, G. S., van Drogen, A., Lee, R., Banerjee, J. J., Hauri, S., O'Reilly, N., Gstaiger, M., Page, R., Peti, W., & Tapon, N. (2019). ASPP proteins discriminate between PP1 catalytic subunits through their SH3 domain and the PP1 C-tail. *Nature Communications*, *10*(1), 1–19. https://doi.org/10.1038/s41467-019-08686-0

Bieging, K. T., Mello, S. S., & Attardi, L. D. (2014). Unravelling mechanisms of p53-mediated tumour suppression. *Nature Reviews Cancer*, *14*(5), 359–370. https://doi.org/10.1038/nrc3711

Bollen, M., Peti, W., Ragusa, M. J., & Beullens, M. (2010). The extended PP1 toolkit: Designed to create specificity. *Trends in Biochemical Sciences*, *35*(8), 450–458. https://doi.org/10.1016/j.tibs.2010.03.002

Brautigan, D. L., & Shenolikar, S. (2018). Protein Serine/Threonine Phosphatases: Keys to Unlocking Regulators and Substrates. *Annual Review of Biochemistry*, *87*, 921–964. https://doi.org/10.1146/annurev-biochem-062917-012332

Cao, L., Huang, Q., He, J., Lu, J., & Xiong, Y. (2013). Elevated expression of iASPP correlates with poor prognosis and chemoresistance/radioresistance in FIGO Ib1-IIa squamous cell cervical cancer. *Cell and Tissue Research*, *352*(2), 361–369. https://doi.org/10.1007/s00441-013-1569-y

Choy, M. S., Hieke, M., Kumar, G. S., Lewis, G. R., Gonzalez-DeWhitt, K. R., Kessler, R. P., Stein, B. J., Hessenberger, M., Nairn, A. C., Peti, W., & Page, R. (2014). Understanding the antagonism of retinoblastoma protein dephosphorylation by PNUTS provides insights into the PP1 regulatory code. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(11), 4097–4102. https://doi.org/10.1073/pnas.1317395111

Choy, M. S., Page, R., & Peti, W. (2012). Regulation of protein phosphatase 1 by intrinsically disordered proteins. *Biochemical Society Transactions*, *40*(5), 969–974. https://doi.org/10.1042/BST20120094

Dancheck, B., Ragusa, M. J., Allaire, M., Nairn, A. C., Page, R., & Peti, W. (2011). Molecular investigations of the structure and function of the protein phosphatase 1-spinophilin-inhibitor 2 heterotrimeric complex. *Biochemistry*, *50*(7), 1238–1246. https://doi.org/10.1021/bi101774g

Daniel Aceytuno, R., Piett, C. G., Havali-Shahriari, Z., Edwards, R. A., Rey, M., Ye, R., Javed, F., Fang, S., Mani, R., Weinfeld, M., Hammel, M., Tainer, J. A., Schriemer, D. C., Lees-Miller, S. P., & Mark Glover, J. N. (2017). Structural and functional characterization of the PNKP-XRCC4-LigIV DNA repair complex. *Nucleic Acids Research*, *45*(10), 6238–6251. https://doi.org/10.1093/nar/gkx275

Dohadwala, M., Da Cruz Silva, E. E. F., Hall, F. L., Williams, R. T., Carbonaro-Hall, D. A., Nairn, A. C., Greengard, P., & Berndt, N. (1994). Phosphorylation and inactivation of protein phosphatase 1 by cyclin- dependent kinases. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(14), 6408–6412. https://doi.org/10.1073/pnas.91.14.6408

Espanel, X., & Sudol, M. (2001). Yes-associated Protein and p53-binding Protein-2 Interact through Their WW and SH3 Domains. *Journal of Biological Chemistry*, *276*(17), 14514–14523. https://doi.org/10.1074/jbc.M008568200

Feng, S., Chen, J. K., Yu, H., Simon, J. A., & Schreiber, S. L. (1994). Two binding orientations for peptides to the Src SH3 domain: Development of a general model for SH3-ligand interactions. *Science*, *266*(5188), 1241–1247. https://doi.org/10.1126/science.7526465

Godin-Heymann, N., Wang, Y., Slee, E., & Lu, X. (2013). Phosphorylation of ASPP2 by RAS/MAPK pathway is critical for its full pro-apoptotic function. *PLoS ONE*, *8*(12), 1–10. https://doi.org/10.1371/journal.pone.0082022

Gorina, S., & Pavletich, N. P. (1996). Structure of the p53 tumor suppressor bound to the ankyrin and SH3 domains of 53BP2. *Science*, *274*(5289), 1001–1005. https://doi.org/10.1126/science.274.5289.1001

Grassie, M. E., Moffat, L. D., Walsh, M. P., & MacDonald, J. A. (2011). The myosin phosphatase targeting protein (MYPT) family: A regulated mechanism for achieving substrate specificity of the catalytic subunit of protein phosphatase type 1δ. *Archives of Biochemistry and Biophysics*, *510*(2), 147–159. https://doi.org/10.1016/j.abb.2011.01.018

Helps, N. R., Barker, H. M., Elledge, S. J., & Cohen, P. T. W. (1995). Protein phosphatase 1 interacts with p53BP2, a protein which binds to the tumour suppressor p53. *FEBS Letters*, *377*(3), 295–300. https://doi.org/10.1016/0014-5793(95)01347-4

Hodge, C. D., Ismail, I. H., Edwards, R. A., Hura, G. L., Xiao, A. T., Tainer, J. A., Hendzel, M. J., & Glover, J. N. M. (2016). RNF8 E3 ubiquitin ligase stimulates Ubc13 E2 conjugating activity that is essential for DNA double strand break signaling and BRCA1 tumor suppressor recruitment. *Journal of Biological Chemistry*, *291*(18), 9396–9410. https://doi.org/10.1074/jbc.M116.715698

Hopkins, J. B., Gillilan, R. E., & Skou, S. (2017). BioXTAS RAW: Improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *Journal of Applied Crystallography*, *50*(5), 1545–1553. https://doi.org/10.1107/S1600576717011438

Hurley, T. D., Yang, J., Zhang, L., Goodwin, K. D., Zou, Q., Cortese, M., Dunker, A. K., & DePaoli-Roach, A. A. (2007). Structural basis for regulation of protein phosphatase 1 by inhibitor-2. *Journal of Biological Chemistry*, *282*(39), 28874–28883. https://doi.org/10.1074/jbc.M703472200

Ito, M., Nakano, T., Erdodi, F., & Hartshorne, D. J. (2004). Myosin phosphatase: Structure, regulation and function. *Molecular and Cellular Biochemistry*, *259*(1–2), 197–209. https://doi.org/10.1023/B:MCBI.0000021373.14288.00

Jia, Y., Peng, L., Rao, Q., Xing, H., Huai, L., Yu, P., Chen, Y., Wang, C., Wang, M., Mi, Y., & Wang, J. (2014). Oncogene iASPP enhances self-renewal of hematopoietic stem cells and facilitates their resistance to chemotherapy and irradiation. *FASEB Journal*, *28*(7), 2816–2827. https://doi.org/10.1096/fj.13-244632

Katz, C., Benyamini, H., Rotem, S., Lebendiker, M., Danieli, T., Iosub, A., Refaely, H., Dines, M., Bronner, V., Bravman, T., Shalev, D. E., Rüdiger, S., & Friedler, A. (2008). Molecular basis of the interaction between the antiapoptotic Bcl-2 family proteins and the proapoptotic protein

ASPP2. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(34), 12277–12282. https://doi.org/10.1073/pnas.0711269105

Lees-Miller, S. P., Sakaguchi, K., Ullrich, S. J., Appella, E., & Anderson, C. W. (1992). Human DNA-activated protein kinase phosphorylates serines 15 and 37 in the amino-terminal transactivation domain of human p53. *Molecular and Cellular Biology*, *12*(11), 5041–5049. https://doi.org/10.1128/mcb.12.11.5041-5049.1992

Li, D. W. C., Liu, J. P., Schmid, P. C., Schlosser, R., Feng, H., Liu, W. B., Yan, Q., Gong, L., Sun, S. M., Deng, M., & Liu, Y. (2006). Protein serine/threonine phosphatase-1 dephosphorylates p53 at Ser-15 and Ser-37 to modulate its transcriptional and apoptotic activities. *Oncogene*, *25*(21), 3006–3022. https://doi.org/10.1038/sj.onc.1209334

Li, G., Wang, R., Gao, J., Deng, K., Wei, J., & Wei, Y. (2011). RNA interference-mediated silencing of iASPP induces cell proliferation inhibition and G0/G1 cell cycle arrest in U251 human glioblastoma cells. *Molecular and Cellular Biochemistry*, *350*(1–2), 193–200. https://doi.org/10.1007/s11010-010-0698-9

Li, S., Shi, G., Yuan, H., Zhou, T., Zhang, Q., Zhu, H., & Wang, X. (2012). Abnormal expression pattern of the ASPP family of proteins in human non-small cell lung cancer and regulatory functions on apoptosis through p53 by iASPP. *Oncology Reports*, *28*(1), 133–140. https://doi.org/10.3892/or.2012.1778

Lim, W. A., Richards, F. M., & Fox, R. (1994). Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature*, *372*, 375–379.

LIN, B. L., XIE, D. Y., XIE, S. B., XIE, J. Q., ZHANG, X. H., ZHANG, Y. F., & GAO, Z. L. (2011). Down-regulation of iASPP in human hepatocellular carcinoma cells inhibits cell proliferation and tumor growth. *Neoplasma*, *58*(3), 205–210. https://doi.org/10.4149/neo

Liu, H., Wang, M., Diao, S., Rao, Q., Zhang, X., Xing, H., & Wang, J. (2009). siRNA-mediated down-regulation of iASPP promotes apoptosis induced by etoposide and daunorubicin in leukemia cells expressing wild-type p53. *Leukemia Research*, *33*(9), 1243–1248. https://doi.org/10.1016/j.leukres.2009.02.016

Liu, W. K., Jiang, X. Y., Ren, J. K., & Zhang, Z. X. (2010). Expression pattern of the ASPP family members in endometrial endometrioid adenocarcinoma. *Onkologie*, *33*(10), 500–503. https://doi.org/10.1159/000319692

Liu, Z. J., Lu, X., Zhang, Y., Zhong, S., Gu, S. Z., Zhang, X. B., Yang, X., & Xin, H. M. (2005). Downregulated mRNA expression of ASPP and the hypermethylation of the 5′-untranslated region in cancer cell lines retaining wild-type p53. *FEBS Letters*, *579*(7), 1587–1590. https://doi.org/10.1016/j.febslet.2005.01.069

Liu, Z. J., Zhang, Y., Zhang, X. B., & Yang, X. (2004). Abnormal mRNA expression of ASPP members in leukemia cell lines. *Leukemia*, *18*(4), 880. https://doi.org/10.1038/sj.leu.2403300

Llanos, S., Royer, C., Lu, M., Bergamaschi, D., Lee, W. H., & Lu, X. (2011). Inhibitory member of the apoptosis-stimulating proteins of the p53 family (iASPP) interacts with protein phosphatase 1 via a noncanonical binding motif. *Journal of Biological Chemistry*, *286*(50), 43039–43044. https://doi.org/10.1074/jbc.M111.270751

Lossos, I. S., Natkunam, Y., Levy, R., & Lopez, C. D. (2002). Apoptosis stimulating protein of p53 (ASPP2) expression differs in diffuse large B-cel and follicular center lymphoma: Correlation with clinical outcome. *Leukemia and Lymphoma*, *43*(12), 2309–2317. https://doi.org/10.1080/1042819021000040017

Lu, B., Guo, H., Zhao, J., Wang, C., Wu, G., Pang, M., Tong, X., Bu, F., Liang, A., Hou, S., Fan, X., Dai, J., Wang, H., & Guo, Y. (2010). Increased expression of iASPP, regulated by hepatitis B virus X protein-mediated NF-κB activation, in hepatocellular carcinoma. *Gastroenterology*, *139*(6), 2183–2194. https://doi.org/10.1053/j.gastro.2010.06.049

Lu, M., Breyssens, H., Salter, V., Zhong, S., Hu, Y., Baer, C., Ratnayaka, I., Sullivan, A., Brown, N. R., Endicott, J., Knapp, S., Kessler, B. M., Middleton, M. R., Siebold, C., Jones, E. Y., Sviderskaya, E. V., Cebon, J., John, T., Caballero, O. L., … Lu, X. (2013). Restoring p53 Function in Human Melanoma Cells by Inhibiting MDM2 and Cyclin B1/CDK1-Phosphorylated Nuclear iASPP. *Cancer Cell*, *23*(5), 618–633. https://doi.org/10.1016/j.ccr.2013.03.013

Mori, S., Ito, G., Usami, N., Yoshioka, H., Ueda, Y., Kodama, Y., Takahashi, M., Fong, K. M., Shimokata, K., & Sekido, Y. (2004). p53 Apoptotic Pathway Molecules Are Frequently and Simultaneously Altered in Nonsmall Cell Lung Carcinoma. *Cancer*, *100*(8), 1673–1682. https://doi.org/10.1002/cncr.20164

Mori, T., Okamoto, H., Takahashi, N., Ueda, R., & Okamoto, T. (2000). Aberrant overexpression of 53BP2 mRNA in lung cancer cell lines. *FEBS Letters*, *465*(2–3), 124–128. https://doi.org/10.1016/S0014-5793(99)01726-3

Mui, M. Z., Zhou, Y., Blanchette, P., Chughtai, N., Knight, J. F., Gruosso, T., Papadakis, A. I., Huang, S., Park, M., Gingras, A.-C., & Branton, P. E. (2015). The Human Adenovirus Type 5 E4orf4 Protein Targets Two Phosphatase Regulators of the Hippo Signaling Pathway. *Journal of Virology*, *89*(17), 8855–8870. https://doi.org/10.1128/jvi.03710-14

Naumovski, L., & Cleary, M. L. (1996). The p53-binding protein 53BP2 also interacts with Bc12 and impedes cell cycle progression at G2/M. *Molecular and Cellular Biology*, *16*(7), 3884–3892. https://doi.org/10.1128/mcb.16.7.3884

Patel, S., George, R., Autore, F., Fraternali, F., Ladbury, J. E., & Nikolova, P. V. (2008). Molecular interactions of ASPP1 and ASPP2 with the p53 protein family and the apoptotic promoters PUMA and Bax. *Nucleic Acids Research*, *36*(16), 5139–5151. https://doi.org/10.1093/nar/gkn490

Pelikan, M., Hura, G. L., & Hammel, M. (2009). Structure and flexibility within proteins as identified through small angle X-ray scattering. *General Physiology and Biophysics*, *28*, 174–189. https://doi.org/10.4149/gpb

Pinto, E. M., Musolino, N. R. C., Cescato, V. A. S., Soares, I. C., Wakamatsu, A., De Oliveira, E., Salgado, L. R., & Bronstein, M. D. (2010). IASPP: A novel protein involved in pituitary tumorigenesis? *Frontiers of Hormone Research*, *38*, 70–76. https://doi.org/10.1159/000318496

Ragusa, M. J., Dancheck, B., Critton, D. A., Nairn, A. C., Page, R., & Peti, W. (2010). Spinophilin directs protein phosphatase 1 specificity by blocking substrate binding sites. *Nature Structural and Molecular Biology*, *17*(4), 459–464. https://doi.org/10.1038/nsmb.1786

Rambo, R. P., & Tainer, J. A. (2013). Super-resolution in solution x-ray scattering and its applications to structural systems biology. *Annual Review of Biophysics*, *42*(1), 415–441. https://doi.org/10.1146/annurev-biophys-083012-130301

Robinson, R. A., Lu, X., Jones, E. Y., & Siebold, C. (2008). Biochemical and Structural Studies of ASPP Proteins Reveal Differential Binding to p53, p63, and p73. *Structure*, *16*(2), 259–268. https://doi.org/10.1016/j.str.2007.11.012

Samuels-Lev, Y., O’Connor, D. J., Bergamaschi, D., Trigiante, G., Hsieh, J. K., Zhong, S., Campargue, I., Naumovski, L., Crook, T., & Lu, X. (2001). ASPP proteins specifically stimulate the apoptotic function of p53. *Molecular Cell*, *8*(4), 781–794.

Sgroi, D. C., Teng, S., Robinson, G., LeVangie, R., Hudson, J. R., & Elkahloun, A. G. (1999). In vivo gene expression profile analysis of human breast cancer progression. *Cancer Research*, *59*(22), 5656–5661.

Shi, Y. (2009). Serine/Threonine Phosphatases: Mechanism through Structure. *Cell*, *139*(3), 468–484. https://doi.org/10.1016/j.cell.2009.10.006

Shopik, M. J., Li, L., Luu, H. A., Obeidat, M., Holmes, C. F. B., & Ballermann, B. J. (2013). Multi-directional function of the protein phosphatase 1 regulatory subunit TIMAP. *Biochemical and Biophysical Research Communications*, *435*(4), 567–573. https://doi.org/10.1016/j.bbrc.2013.05.012

Skene-Arnold, T. D., Luu, H. A., Uhrig, R. G., De Wever, V., Nimick, M., Maynes, J., Fong, A., James, M. N. G., Trinkle-Mulcahy, L., Moorhead, G. B., & Holmes, C. F. B. (2013). Molecular mechanisms underlying the interaction of protein phosphatase-1c with ASPP proteins. *Biochemical Journal*, *449*(3), 649–659. https://doi.org/10.1042/BJ20120506

Soubeyrand, S., Schild-Poulter, C., & Haché, R. J. G. (2004). Structured DNA promotes phosphorylation of p53 by DNA-dependent protein kinase at serine 9 and threonine 18. *European Journal of Biochemistry*, *271*(18), 3776–3784. https://doi.org/10.1111/j.1432-1033.2004.04319.x

Terrak, M., Kerff, F., Langsetmo, K., & Tao, T. (2004). Structural basis of protein phosphatase 1 regulation. *Nature*, *429*, 780–784.

Trigiante, G., & Lu, X. (2006). ASPPs and cancer. *Nature Reviews Cancer*, *6*(3), 217–226. https://doi.org/10.1038/nrc1818

Vigneron, A. M., Ludwig, R. L., & Vousden, K. H. (2010). Cytoplasmic ASPP1 inhibits apoptosis through the control of YAP. *Genes and Development*, *24*(21), 2430–2439. https://doi.org/10.1101/gad.1954310

Wang, L., Xing, H., Tian, Z., Peng, L., Li, Y., Tang, K., Rao, Q., Wang, M., & Wang, J. (2012). IASPPsv antagonizes apoptosis induced by chemotherapeutic agents in MCF-7 cells and mouse thymocytes. *Biochemical and Biophysical Research Communications*, *424*(3), 414–420. https://doi.org/10.1016/j.bbrc.2012.06.124

Yang, J. P., Hori, M., Sanda, T., & Okamoto, T. (1999). Identification of a novel inhibitor of nuclear factor-κB, RelA- associated inhibitor. *Journal of Biological Chemistry*, *274*(22), 15662–15670. https://doi.org/10.1074/jbc.274.22.15662

Yang, J. P., Hori, M., Takahashi, N., Kawabe, T., Kato, H., & Okamoto, T. (1999). NF-κB subunit p65 binds to 53BP2 and inhibits cell death induced by 53BP2. *Oncogene*, *18*(37), 5177–5186. https://doi.org/10.1038/sj.onc.1202904

Zhang, B., Xiao, H. J., Chen, J., Tao, X., & Cai, L. H. (2011). Inhibitory member of the apoptosis-stimulating protein of p53 (ASPP) family promotes growth and tumorigenesis in human p53-deficient prostate cancer cells. *Prostate Cancer and Prostatic Diseases*, *14*(3), 219–224. https://doi.org/10.1038/pcan.2011.25

Zhang, X., Wang, M., Zhou, C., Chen, S., & Wang, J. (2005). The expression of iASPP in acute leukemias. *Leukemia Research*, *29*(2), 179–183. https://doi.org/10.1016/j.leukres.2004.07.001

Zhao, J., Wu, G., Bu, F., Lu, B., Liang, A., Cao, L., Tong, X., Lu, X., Wu, M., & Guo, Y. (2010). Epigenetic silence of ankyrin-repeat-containing, SH3-domain-containing, and proline-rich-region-containing protein 1 (ASPP1) and ASPP2 genes promotes tumor growth in hepatitis B virus-positive hepatocellular carcinoma. *Hepatology*, *51*(1), 142–153. https://doi.org/10.1002/hep.23247

Zhuo, S., Clemens, J. C., Hakes, D. J., Barford, D., & Dixon, J. E. (1993). Expression, purification, crystallization, and biochemical characterization of a recombinant protein phosphatase. *Journal of Biological Chemistry*, *268*(24), 17754–17761. https://doi.org/10.1016/s0021-9258(17)46769-x