ABSTRACT

Title of Dissertation: STRUCTURAL STUDIES OF NEP1/EMG1 RNA METHYLTRANSFERASE AND ATYPICAL RIO3 KINASE

Seth Thomas, Doctor of Philosophy, 2011

Dissertation Directed By: Assistant Professor Nicole LaRonde-LeBlanc
Department of Chemistry and Biochemistry

Nucleolar Essential Protein 1 (Nep1) is required for small subunit (SSU) ribosomal RNA (rRNA) maturation and is mutated in Bowen–Conradi Syndrome. Although yeast (*Saccharomyces cerevisiae*) Nep1 interacts with a consensus sequence found in three regions of the SSU rRNA, the molecular details of the interaction are unknown. Nep1 is a SPOUT RNA methyltransferase, and can catalyze methylation at the N1 position of pseudouridine. Nep1 is also involved in assembly of Rps19, an SSU ribosomal protein, into the SSU. Mutations in Nep1 that result in decreased methyl donor binding do not result in lethality, suggesting that enzymatic activity may not be required for function, and RNA binding may play a more important role. To study these interactions, the crystal structures of the ScNep1 dimer and its complexes with RNA were determined. The results demonstrate that Nep1 recognizes its RNA site via base-specific interactions and stabilizes a stem-loop in the bound RNA. Furthermore, the observed RNA structure contradicts the structures of the Nep1-binding sites within mature rRNA, suggesting that the Nep1 changes rRNA

structure upon binding. Finally, a uridine base is bound in the active site of Nep1, positioned for a methyltransfer at the C5 position, supporting Nep1's role as an N1-specific pseudouridine methyltransferase.

In addition to the work completed with the Nep1 project, structural characterization of Rio3 Kinase is reported, as well as collaborative work in the structure determination of Ubiquitin and Ubiquitin complexes and Iodotyrosine Deiodinase.

STRUCTURAL STUDIES OF NEP1/EMG1 RNA METHYLTRANSFERASE AND ATYPICAL RIO3 KINASE

By

Seth Thomas

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2011

Advisory Committee:

Professor Jason Kahn, Chair
Professor Nicole LaRonde-LeBlanc
Professor George Lorimer
Professor David Fushman
Professor John Moult, Dean's Representative

© Copyright by
Seth Thomas
2011

Acknowledgements

I would first like to thank my advisor, Dr. Nicole LaRonde-LeBlanc for her support, guidance, advice, and patience with me during the last five years. I can easily remember as if it was yesterday, walking into the lab and seeing empty, black bench tops. Now, five years later, I am unable to express enough appreciation for her time and persistence, without which, this dissertation would not be possible.

I would also like to thank all past and present members of the LaRonde Lab, especially Irene Kiburu, Chris Keller, and Dr. Aga Szyk for their daily conversations, help and support.

Thank you also to all the members of my review committee. Dr. Kahn for being the chair of my committee and providing leadership. Dr. Lorimer for always having your door open and providing support and direction to the Center for Biomolecular Structure and Organization. Extended thanks to Dr. Fushman for not only being a member of my review committee but also providing me the opportunity to work and teach with members of his laboratory. Dr. Moult for dedicating his time as the Dean's representative on the review committee.

Thanks to the ARCS Foundation for providing support that allowed me to focus on research.

Thanks to Dr. Patrick McTamney, Jennifer Adler and Dr. Steven Rokita for our successful collaboration with IYD and I wish everyone in the Rokita Lab success in the future. Working on IYD was an experience I will not forget for the rest of my life. I am a better all around scientist having worked with the Rokita Lab.

I would like to extend my heartfelt gratitude to Mrs. Paulette Frazier. She is an extraordinary person that put up with me for five years. Without Paulette, I would have been completely lost at the University of Maryland. Simply put, she is amazing.

Thank you to all the members of my family: my mom and dad, Dean, Belinda, and Nikki for their continual support and frequent motivational encouragement.

Finally, I would like to thank my future wife Misty. From long, long distance to just long distance to our upcoming wedding, you have provided the support, motivation, and love I have needed. I love you and cannot wait to be your husband!

Table of Contents

List of Tables

List of Figures

Abbreviations

Å	Angstrom
Af	*Archaeoglobis fulgidus*
Ala	Alanine
APS	Advanced Photon Source
Arg	Arginine
Asn	Asparagine
Asp	Aspartic Acid
ATP	Adenosine-5'-triphosphate
BIS-TRIS	Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane
BME	β-mercaptoethanol
CCD	Charged Coupled Device
CCP4	Collaborative Computational Project No. 4
COOT	Crystallographic Object Oriented Toolkit
Cys	Cysteine
Da	Dalton
ºC	Degrees Celsius
DIT	Di-iodotyrosine
DNA	Deoxyribonucleic Acid
EDTA	Ethylenediaminetetraacetic Acid
EM	Electron Microscopy
ePK	Eukaryotic Protein Kinase
FMN	Flavin Mononucleotide
Gln	Glutamine
Glu	Glutamic Acid
H	Hours
HEPES	4-(2-hydroxyethyl)-1-pipeazineethanesulfonic Acid
HPLC	High-Pressure Liquid Chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
IYD	Iodotyrosine Deiodinase
K	Kelvin
LB	Luria Broth
LC MS/MS	Liquid Chromatography Mass Spectrometry/Mass Spectrometry
Leu	Leucine
M	Molar
MES	2-(N-morpholino)ethanesulfonic Acid
Mg	Milligram
MIT	Mono-iodotyrosine
mL	Milliliter
mM	Millimolar
MW	Molecular Weight
NEP1	Nucleolar Essential Protein 1
NECAT	Northeastern Collaborative Access Team
NMR	Nuclear Magnetic Resonance

PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
PHENIX	Python-based Hierarchical Environment for Integrated Xtallography
RMSD	Root Mean Square Deviation
RNA	Ribonucleic Acid
RPM	Revolutions Per Minute
rRNA	Ribosomal Ribonucleic Acid
S	Svedburg Unit
SAD	Single Wavelength Anomalous Diffraction
SAH	S-adenosylhomocysteine
SAM	S-adenosyl Methionine
Sc	*Saccharomyces cerevisiae*
SDS-PAGE	Sodium Dodecyl Sulface Polyacrylamide Gel Electrophoresis
Se-Met	Selenium Methionine
Ser	Serine
SER	Surface Entropy Reduction
SPOUT	SpoU and TrmD Family Name
SSU	Small Ribosomal Subunit
TBE	Tris/Borate/EDTA
TEV	Tobacco Etch Virus
Thr	Threonine
Tris	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet
Val	Valine
wHTH	Winged Helix-Turn-Helix
µL	Microliter
µM	Micromolar

Chapter 1: Introduction

1.1: Overview

Biochemistry is known as the chemistry of life. It is the study of structure-function relationships to determine cellular processes. Primary DNA sequences are transcribed into RNA, and then later translated into amino acids by the ribosome. The twenty possible natural amino acids are connected linearly by a peptide bond, forming what is known as a polypeptide chain. These polypeptide chains then take on a secondary structure, specifically determined by the sequence of amino acids in the chain, leading to the formation of alpha helices and beta sheets. The final proteins products are globular arrangements of secondary structure (also known as tertiary structure) stabilized into a low energy state. Once folded correctly into proper three-dimensional globular structures, proteins are tasked to carry out varied cellular processes essential to life of all organisms.

At the core of protein biochemistry is the synthesis of the polypeptide chain. A ribonucleoprotein complex called the ribosome performs this process via a mechanism known as translation; this process translates the genetic code into amino acids leading to the synthesis of a fully functional protein. The eukaryotic ribosome is made up of ribosomal RNA (rRNA) and ribosomal proteins assembled into two mature subunits, one larger than the other. The large subunit (60S) and small subunit (40S) come together independently during translation to form the complete 80S ribosome. It takes a highly ordered, well-regulated scheme for the cell to manufacture rRNA and process it into its mature fragments. In eukaryotes, the initial

35S pre-rRNA (part of the 90S pre-ribosome) is chopped into smaller pieces through a series of cleavages (*1-3*). The final mature rRNA products are the 5.8S, 25S and the 18S (Fig. 1.1). There are estimated to be over 170 proteins that are involved in the maturation of the eukaryotic ribosome (*3*). These proteins carry out multiple different functions toward the final goal of producing a fully functional mature 80S ribosome, including direction of cleavage steps, rRNA modifications (such as methylation and pseudouridylation), transport, and chaperone function (3-9). The structural characterization of known ribosomal processing factors with unknown functions can provide a molecular analysis into their role in ribosome biogenesis.

Currently, the only two ways to acquire molecular resolution better than 4.0 Å of protein structure is by x-ray crystallography or nuclear magnetic resonance (NMR). In the late 1950's, myoglobin was the first protein for which a crystal structure was solved, credited to John Kendrew with a Nobel Prize in 1962 (*10*). Including the contribution from NMR, which saw its first protein structure in 1984 by Kurt Wüthrich, there are greater than 65,675 protein structures deposited to the protein database (*11, 12*). Amazingly, just five short years ago, there were fewer than 38,000 submitted structures (12). X-ray crystallography methods have advanced so quickly the past half-decade that it is an arduous task to keep up with new technology. The widespread availability/affordability of high-throughput machinery, in house diffraction systems, and improved/new crystallographic software (Phenix/COOT/CCP4) are major factors for this recent success. Even with all the recent success, x-ray crystallography historically and currently has the same rate-limiting step, and that is the growth of suitable crystals that yield high-resolution

diffraction patterns. This manuscript will describe multiple projects, both successful and challenging, and will provide a glimpse into the exciting trials of x-ray crystallography.

1.2: Background: Nep1

Nucleolar essential protein 1 (Nep1) is a highly conserved protein required for ribosome biogenesis and found in organisms from archaea to humans. Nep1 was first identified in fission yeast (*Schizosaccharomyces pombe*) as a high-copy number suppressor of a mating deficiency caused by a mutation in the effector domain of Ras1 and thus named Mra1 (multicopy suppressor of Ras 1) (13). In yeast, Nep1 has also been designated as Emg1, for essential for mitotic growth 1 (14). Emg1 is required for 18S SSU rRNA maturation (*14, 15*). To release the yeast 18S SSU rRNA from the pre-rRNA, cleavage must occur at four sites, designated sites A0, A1, A2 and D (*2, 3*). Cleavage at site A0 and A1 truncates the 5'-end of the rRNA transcript, while cleavage at A2 results in separation of a 20S pre-rRNA species from the 5.8S and 25S. Final cleavage at site D produces the mature 18S SSU rRNA (Fig. 1.1). Depletion of Emg1 results in loss of cleavage at site A2, which results in an accumulation of a 21S SSU rRNA species (15). This is a result of cleavage at site A3, which normally happens after site A2 cleavage in the maturation of the 5'-end of the 5.8S rRNA species (*2*).

1.3: Introduction to the SPOUT Family

Recently in 2008, work published by Taylor *et al*. and Leulliot *et al*., provided the first looks at Nep1/Emg1 with a dimeric structure from archaebacterium

Methanocaldococcus jannaschii and a monomeric structure from *Saccharomyces cerevisiae* (*16, 17*). Structural analysis of archaeal Nep1, of our own then unpublished work and the work by Taylor *et al*, revealed structural homology to the SPOUT family of RNA methyltransferases. The name SPOUT is derived from the family's original methyltransferase members, SpoU and TrmD (*18*). The family is now much larger, but they all still have shared structural characteristics (*16, 18, 19*). The members of the family each contain a SPOUT methyltransferase domain and may contain additional inserted domains required for selectivity of RNA substrates. Members of the SPOUT family all contain a signature methyltransferase knotted fold, where S-adenosylmethionine (SAM) is the source for the methyl group used in the methyl transfer. Thus far, SPOUT methyltransferases have been implicated in the methylation of ribose 2'-OH (2'-O-ribose), guanine N1 (m1G), uridine C3 (m3U) and N3 of pseudouridine (m3ψ) (*19, 20*). Within the family, Nep1 proteins are most closely related to the TrmD and RsmE subfamilies, which catalyze m1G and m3U modifications, respectively (*19*). All SPOUT methyltransferases characterized thus far are dimeric, as is Nep1.

1.4: Nep1 is a Pseudouridine RNA Methyltransferase

RNA-binding screens have indicated that Nep1 binds to the consensus sequence C/UUCAAC (*21*). This sequence is found in three regions of 18S rRNA in yeast including the stem loop of helix 31, bases 1189-1193, and part of the stem of helix 42, bases 1565–1570 (Fig. 1.2). In the RNA three-hybrid screen that identified this consensus sequence, all the binding sites were located in the unpaired region of a stem loop structure predicted by RNA secondary structure predictions (*21*).

Recent work has identified Nep1 as a pseudouridine N1-methyltransferase (*22*) (Fig. 1.3). Wurm *et al*. showed that both human and *M*. *jannaschii* Nep1 are capable of catalyzing methylation on pseudouridine-containing RNA sequences, especially C/UψCAAC. This sequence is the Nep1 site found in the stem loop of helix 31 of yeast 18S rRNA mentioned above (*22*). This pseuoduridine, located at position 1189, is hypermodified in yeast, with a 3-amino-3-carboxypropyl group on the N3 position and a methyl group on the N1 position (m1acp3ψ) (*23*). Nep1 is the first identified example of an N1-specific pseudouridine methyltransferase (*22*).

1.5: Nep1 Interactions with Rps19 and SnR57

Nep1 plays an unclear role in the loading of Rps19 into the SSU. Rps19 is a small ribosomal subunit (SSU) ribosomal protein found only in archaeal and eukaryotic ribosomes and is also known as a protein mutated in roughly 25% of cases of Diamond–Blackfan Anemia, a rare congenital disease characterized by defects in red blood cell development and predisposition to cancers (*24, 25*). Rps19 is an essential protein, and its depletion also results in the loss of cleavage at site A2 (*16, 25-28*). In yeast, studies performed by Buchhaupt *et al*., overexpression of Rps19 was shown to partially suppress the phenotype observed with loss of Nep1 function. Surprisingly, this suppression is enhanced by deleting snr57, the C/D box snoRNA that guides 2'-O-ribose methylation at 18S G1572 (*21, 29*). In fact, deletion of Nep1 is not lethal in an snR57 deletion mutant (*21*). Taken together, it appears that Nep1 aids in Rps19 loading and is essential in the presence of snr57. This function is either through direct interaction with Rps19, or alteration of the rRNA structure (*21*). Since the defects in rRNA processing observed in Nep1 and Rps19 depletion are identical,

it is likely that some of the ribosome biogenesis defect caused by loss of Nep1 is a consequence of improper Rps19 assembly.

As mentioned above and shown in Figure 1.2, Nep1 consensus binding sequence C/UUCAAC is found in helix 31 (1188-1193 and helix 42 (1565-1570). Both these binding sites are near helix 41, where Rps19 was shown to bind using cryo electron microscopy (cryo-EM) reconstructions (*21, 30*). In addition, one of the sites, between bases 1565 and 1570, overlaps with the predicted binding site of snR57 (bases 1569–1583).

1.6: Nep1 Objectives

The initial focus of my research was to determine the first structure of Nep1 and use structural analysis to understand its role in ribosome biogenesis. When, as mentioned, Taylor *et al*. reported the first Nep1 structure in 2008 (*16*) my research took on a new direction. Utilizing the Nep1 structure, we asked the following questions:

1. Is it possible to crystallize and solve the structure of Nep1 in complex with cognate target RNA,
2. If so, what drives specificity for the target RNA, and what is a plausible mechanism for RNA methylation.
3. What is the functional role of Nep1 in ribosome biogenesis?

To address these questions, in chapter 2 I will describe the purification and initial crystallization and structure solution of *Archaeaglobis fulgidus* Nep1 (AfNep1) as well as the structure of *Saccharomyces cerevisiae* Nep1/Emg1 (ScNep1) solved via

seleno-methionine substitution and molecular replacement, respectively. Chapter 2 will also describe the methods utilized, including RNA oligonucleotide design, to acquire two x-ray structures of a complex formed between ScNep1 and a 14 base cognate RNA oligonucleotide. Chapters 3 and 4 will introduce all four structures determined from the experiments detailed in chapter 2 and provide structural analysis highlighting the active site, dimer interface and interactions with bound RNA. Conclusions to the Nep1 project will be presented in chapter 5.

1.7: Rio3 Background

Kinases are involved in a wide range of cellular processes including and not limited to cell cycle regulation, DNA replication, transcription, and ribosome biogenesis (*31, 32*). Atypical protein kinases are defined as members of the kinase superfamily that possess phosphoryl transfer activity without significant sequence homology to canonical eukaryotic protein kinases (ePK). The RIO kinase family is classified within this group of atypical kinases because of the very low sequence homology to ePKs and the ability to phosphorylate serine residues (*33, 34*). Structural studies of Rio1 and Rio2 show the RIO domain is structurally similar to ePK's kinase domain (*35-37*). The founding member of the RIO family (Rio1) has been shown along with Rio2 to be essential in ribosome biogenesis, more specifically the maturation of the 20S to the fully mature 18S ribosome (*38-41*).

There are over 500 serine/threonine/tyrosine kinases in the human genome (*42*). Of those kinases, there are 13 families of atypical kinases, which represent 40 of the 500+ kinases (*42*). The functional subdomains of the 250-300 amino acid kinase domain are structurally conserved among the canonical eukaryotic protein

kinases (*43*). These subdomains include: nucleotide binding loop, known as the P-loop; hinge region, which interacts with the adenine of adenosine triphosphate; "catalytic loop" containing catalytic residues asparagine and aspartate; activation loop or APE; metal binding loop; and peptide binding sub-domains on the C-terminus (*39, 43-45*). Structural studies of archaeal Rio1 and Rio2 have shown the absence of the activation loop sub-domain and C-terminal peptide binding sub-domains (*35, 37*).

All metazoa carry the gene for the third member of the RIO kinase family. Rio3 is conserved in all multi-cellular eukaryotes. However, Rio3 is not present in yeast and archaea, unlike Rio1 and Rio2. Even though Rio1 and Rio2 share conservation through archaea and yeast, Rio3 and Rio1 are more closely related. With an overall sequence alignment (ClustalW) identity of 45% for the RIO kinase domain alone, Rio3 human (RioK3) and Rio1 human (RioK1) are highly homologous and come from the same phylogeny (Fig. 1.4A) (*46, 47*). Rio2 human (RioK2) and RioK3 are less related, with an overall alignment identity for the RIO domain identity of just 19%. Each member has a unique N-terminal and C-terminal domain/sub-domain. Rio2's conserved N-terminal winged helix turn helix (wHTH) domain is known via archaeal Rio2 structures. However, eukaryotic Rio1 has an extended N-terminal domain of unknown structure not present in archaea. Rio3 contains a longer unknown domain (in comparison to Rio1) on the N-terminus, consisting of ~200 amino acids (*48*). Figure 1.4B shows the domain arrangement for the three members of the RIO family.

Recently, RioK3 has been identified as an oncogene in pancreatic ductal adenocarcinoma and was shown to significantly alter cytoskeletal architecture when overexpressed (*49*). The characteristic phenotype is the formation of lamellipodia, an appendage like protrusion commonly seen in migrating and invasive cells (*50*). Overexpression of RioK3 was also shown to promote this invasive cell phenotype by the activation of a small GTPase known as Rac (*49*). It is still uncertain as to how RioK3 activates Rac and whether RioK3 requires kinase function, either through autophosphorylation or through phosphorylation of a currently unknown substrate. The significant uncertainties of RioK3 combined with its characterization as an oncogene make RioK3 a prime target for further structural and biochemical studies.

1.8: Rio3 Studies

A major portion of my tenure at the University of Maryland has been working on the purification and crystallization of Rio3. There have been successes in purification protocols that will be detailed in chapter 6. For every success in the Rio3 project, there has been an equal roadblock at protein stability, initial crystal screening, as well as in the optimization of diffraction from acquired crystals. Chapter 7 will provide details to the exhaustive approach to improve crystal quality and diffraction. Despite my inability to produce suitable crystals prior to the end of my time here, the work described will lay the foundation for future success in the structural and biochemical characterization of Rio3.

1.9: Additional Studies

In addition to the previously described work, I have participated in a variety of collaborative projects. Chapter 8 will describe work performed with Dr. David Fushman, a co-member of the Center for Biomolecular Structure and Organization, which focused on Ubiquitin and Ubiquitin in complex with E2-25K. Finally, chapter 9 contains the published work I completed in collaboration with then senior graduate student Dr. Patrick McTamney and his advisor, Dr Steven Rokita, at the University of Maryland. The collaborations were exciting and provided experiences that guided me to life lessons in protein crystallography and collaboration formation.

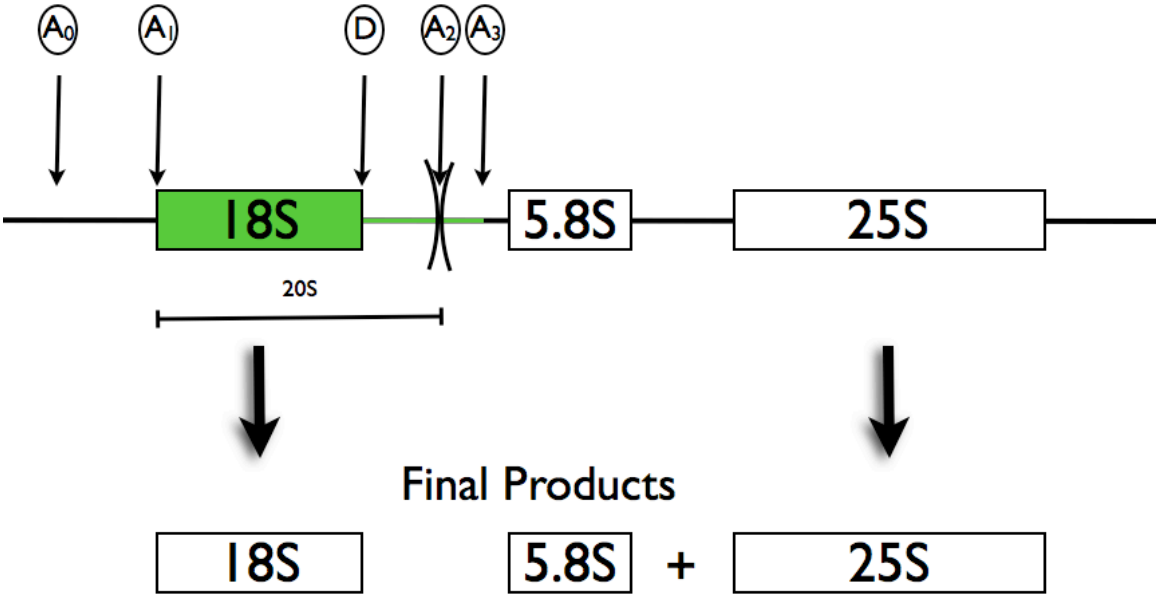


Figure 1.1: 35S Pre-rRNA Processing

Thin arrows indicate specific cleavage sites. Cleavage at A_0 and A_1 occur first off the 5'-end. Next, cleavage at site A_2 results in the separation of the 20S pre-rRNA, which later becomes the final mature 18S rRNA. Depletion of Nep1 (loss of cleavage at site A_2) results in accumulation of the 21S pre-rRNA portion depicted in green, caused by the continuation of normal cleavage at site A_3.

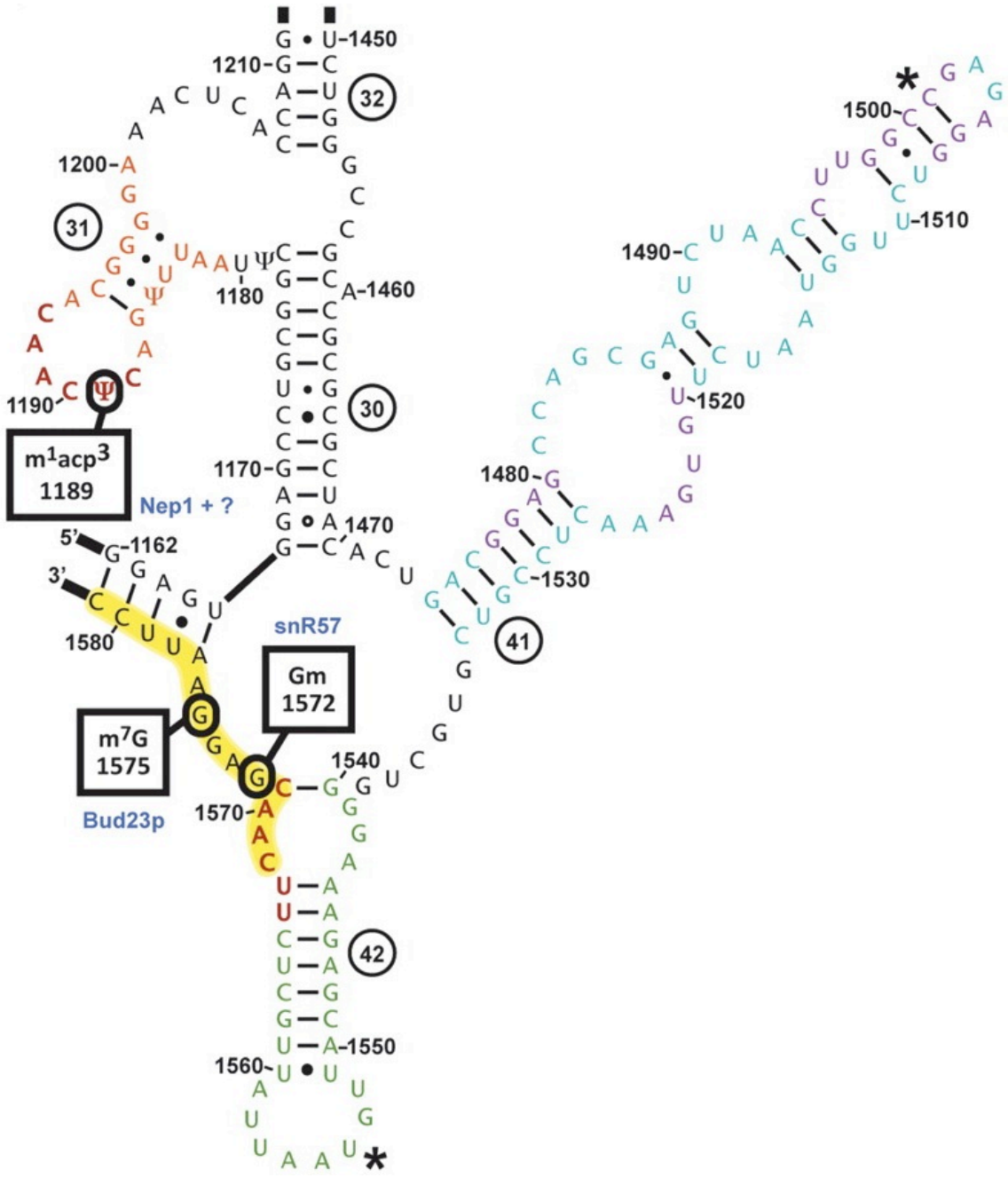


Figure 1.2: 18S RNA Sequence and Structure

Based on the secondary structure prediction for S. cerevisiae 18S rRNA from the Comparative RNA Web Site (www.rna.ccbb.utexas.edu). Selected regions of secondary structure are colored orange, cyan and green for helix 31, 41 and 42 respectively. Red color indicates Nep1 consensus sites. The snR57-binding site is

highlighted in yellow. Purple denotes parts of rRNA within 4 Å of S19 (shown as blue cartoon). Asterisk indicates the site of interaction between helix 41 and 42. Small circles around bases represent bases with post-translational methylation modification (*51*).

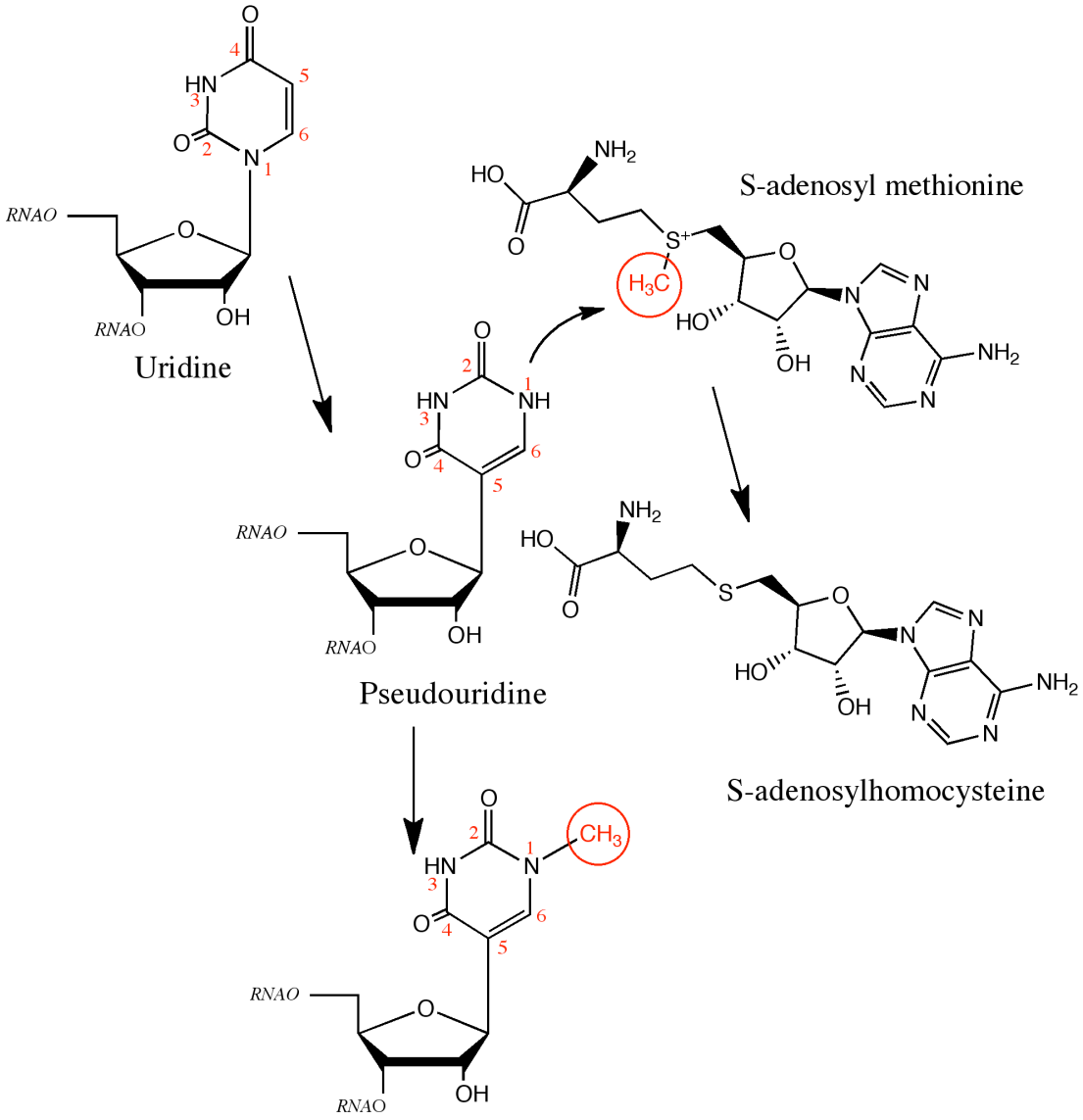


Figure 1.3: Nep1 Catalyzed Reaction

Nep1 catalyzes the transfer of the methyl group (circled in red) from S-adenosyl methionine (SAM) to the N1 position of pseudouridine (ψ) resulting in N1-methyl psuedouridine and left over S-adenosylhomocysteine (SAH).

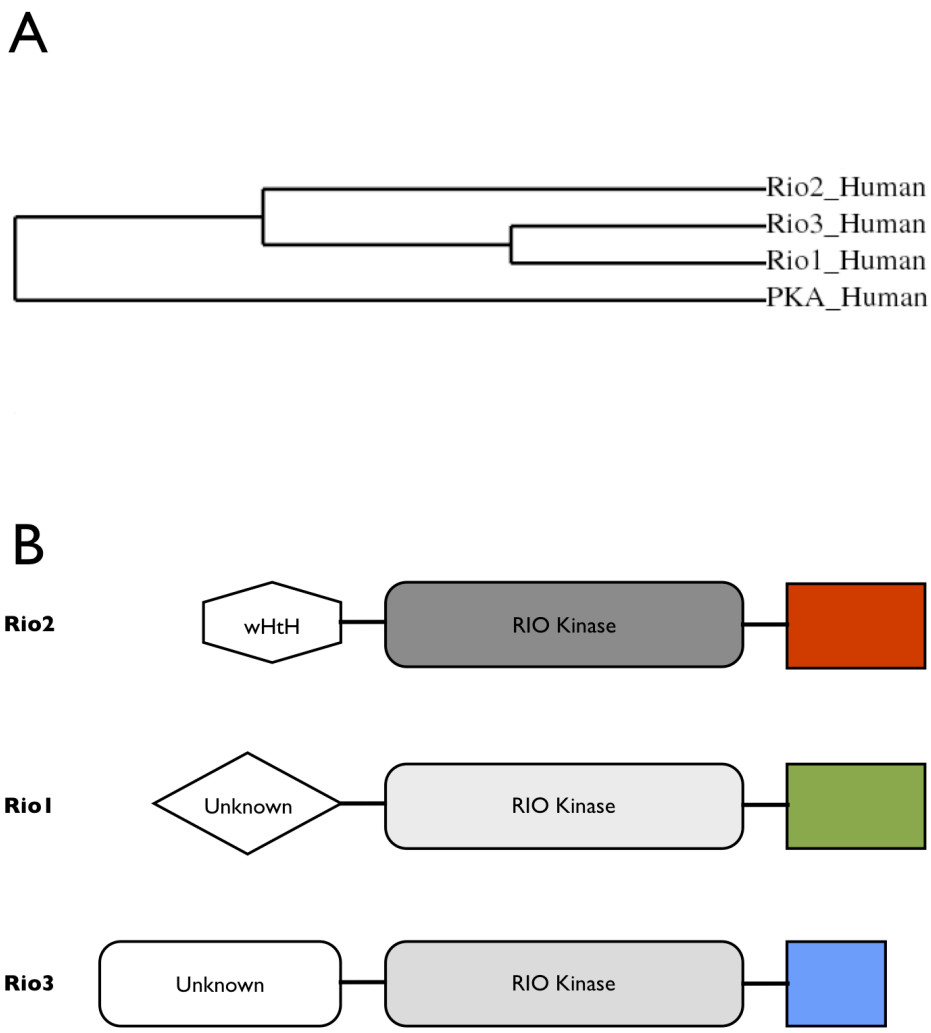


Figure 1.4: Eukaryotic Rio Family Cladogram and Domain Arrangement

A) Cladogram showing evolutionary relationships (*47*). PKA (c-AMP dependent protein kinase A) represents canonical eukaryotic kinase. **B**) Rio family members include a conserved atypical Rio kinase domain, with Rio1 and Rio3 being more similar (light grey used for Rio1 and Rio3 to show similarity vs dark grey for Rio2). Each family member has a unique N-terminal domain (Rio1 and Rio3 both unknown folds; Rio3 predicted winged helix turn helix (wHtH) motif and unique C-terminal residues.

Chapter 2: Purification and Crystallography of AfNep1 and ScNep1

2.1: Purification Protocol for AfNep1 and ScNep1

Archaeoglobis fulgidus was originally discovered in 1987 deep in an Italian thermal vent (*52, 53*). This hyperthermophilic organism was later discovered to optimally thrive at 76°C (*54*). In order to survive, *A. fulgidus* has evolved to prevent the denaturing of proteins at high temperatures. The elevated thermal stability of *A. fulgidus* proteins allows the design of a purification protocol for their expression in *E. coli* that utilizes wide differences in thermal denaturation properties.

The full-length Nep1 gene was PCR amplified from *A. fulgidus* and *S. cerevisiae* genomic DNA (ATCC) and subcloned into a plasmid vector (pDest527) containing an N-terminal 6X histidine tag followed by a Tobacco Etch Virus (TEV) protease cleavage site (plasmids obtained from Protein Expression Laboratory, SAIC, Frederick, MD) (*55*). These constructs were transformed into *E. coli* Rosetta™ -DE3-pLysS cells (Novagen). Expression cultures were grown in LB media at 37 ºC and induced with 1 mM isopropyl-beta-D-thiogalactopyranoside (IPTG) at OD600 of 0.6 for 4 hours for AfNep1. For ScNep1, cultures were grown at 37°C until OD600 of 0.6, then cooled to 20°C and induced with 1 mM IPTG for overnight expression. After expression, the cells were harvested by centrifugation. For AfNep1, cells were resuspended in 50 mL of 50 mM Tris pH 8.0, 300 mM NaCl, 0.2 % β-mercaptoethanol, 0.2 X Bugbuster™ (Novagen) and 0.1 mg/ml DNase1 (Roche) per liter of expression culture and stirred at room temperature. After 30 minutes, the lysate was transferred to centrifuge tubes and placed in a 75 °C water bath for 15 minutes to denature the *E. coli* proteins and then immediately centrifuged for 20

minutes at 18000 rpm (rotor: Beckman 45 Ti) in an ultracentrifuge to remove insoluble/denatured material. For ScNep1, cells were resuspended in 50 mL of 25 mM Tris pH 8.0, 200 mM NaCl, 0.2% β-mercaptoethanol, 0.2 X Bugbuster™ and 0.1 mg/ml DNase1 (Roche) per liter of expression culture and stirred at 4°C. After 45 minutes, the lysate was transferred to centrifuge tubes and centrifuged for 20 minutes at 18000 rpm (45 Ti). The supernatants were passed through a 0.22 μM filter and loaded onto a 5 ml HisTrap HP column (GE Healthcare) equilibrated in buffer containing 50 mM Tris pH 8.0, 300 mM NaCl and 0.2 % β-mercaptoethanol for Nep1 and 25 mM Tris pH 8.0, 200 mM NaCl and 0.2% β-mercaptoethanol for ScNep1. The bound AfNep1/ScNep1 was washed with 100 mM imidazole in the respective equilibration buffer and then eluted with a gradient from 100 mM to 1 M imidazole in equilibration buffer in 20 column volumes. The fractions containing AfNep1/ScNep1 were pooled and dialysed overnight into 25 mM Tris pH 8.0, 300 mM NaCl, 1 mM EDTA and 0.2 % β-mercaptoethanol for AfNep1 and 25 mM Tris pH 8.0, 200 mM NaCl, 1mM EDTA and 0.2 % β-mercaptoethanol for ScNep1. In the case of AfNep1, attempts made to remove the tags by TEV protease cleavage were unsuccessful. The tag was cleaved in ScNep1. The resulting protein contained a single extra glycine residue at the N-terminus. The cleaved ScNep1 protein was passed over a 5 ml HisTrap HP column (GE Healthcare) a second time to remove uncleaved protein and TEV protease (*55*). The proteins were concentrated to 5 - 10 mg/ml and further purified to >99% by size exclusion chromatography using a Superdex 75™ column (GE Healthcare) equilibrated in 20 mM Tris pH 8.0, 300 mM NaCl, 1 mM EDTA and 0.2% β-mercaptoethanol for Nep1, and a Superdex 200™ column (GE Healthcare)

equilibrated in 20 mM Tris pH 8.0, 200 mM NaCl, 1 mM EDTA and 0.2% β-mercaptoethanol for ScNep1. The protein was concentrated again to 8 mg/ml for AfNep1 and 20 mg/ml for ScNep1. This preparation was stored at 4°C until crystallization screening. Se-Met AfNep1 was expressed in minimal media with all amino acids supplemented except for methionine, which was replaced by seleno-methionine (Se-Met). Reducing agent (0.2% β-mercaptoethanol) was added to Se-Met protein immediately prior to crystallization screening.

2.2: Crystallography

2.2A: RNA Oligonucleotide Design

Table 2.1 shows the RNA oligonucleotides of varying lengths and sequences purchased HPLC-purified from IDT. At the time of oligonucleotide design, the only information available was the identification of the ScNep1 consensus binding site, C/UUCAAC, and RNA secondary structure prediction of the target sequence (*21, 51*). It was not known that pseudouridine was required for methylation by ScNep1. The initial design utilized different lengths of the identified sequence (oligo #1, #3 and #4) and extending the consensus sequence with Watson-Crick G-C base pairs to stabilize the predicted stem-loop secondary structure (oligo #2).

ScNep1 was mixed at a 1:1 molar ratio with the RNA oligonucleotides resuspended in 10 mM Tris pH 8.0. The mixtures were analyzed by gel electrophoresis on a 6% TBE-polyacrylamide gel and visualized by UV after ethidium bromide staining to determine stoichiometric complex formation. The 14-base RNA oligo #2 containing the sequence 5'-GGGCUUCAACGCCC-3' was the

best at producing stoichiometric complexes. Oligo #3 containing 5'-CUCUUCAACGAG-3' also showed good complex formation, however no crystals were obtained with RNA present in the active site.

2.2B: Crystallization Conditions

Initial crystallization conditions were obtained through utilization of several sparse matrix screens (Qiagen, Emerald Biosystems Inc., Nextal, Hampton Research) with the sitting drop vapor diffusion method. Crystals of the AfNep1 protein were obtained from drops containing 3-4 M sodium chloride, 100 mM HEPES buffer pH 7.0 – 8.0 placed over 1 mL reservoirs. Crystals of free ScNep1 were obtained from several conditions, including 50 mM BIS-TRIS, pH 6.5, 50 mM ammonium sulfate, 30% pentaerythritol ethoxylate (15/4 EO/OH) and 20% glycerol, used for data collection. The ScNep1/1RNA complex crystallized in 100 mM HEPES, pH7.0 and 10% PEG (polyethylene glycol) 6000 and crystals of the ScNep1/2RNA complex were obtained in 50 mM MES, pH 6.0, 15 mM magnesium sulfate and 7% PEG 4000. Diffraction quality crystals were obtained after 4-6 days at 20 °C. (Fig. 2.1)

2.2C: Data Collection

Crystals were flash frozen in mother liquor supplemented with 20% glycerol for all crystals. Cryogenic conditions are used to protect crystals from radiation damage during data collection as well as to prevent icing during cryo-preservation. Diffraction data for the native and Se-Met were collected at 100 K at the NECAT beamline at the Advanced Photon Source (APS), Argonne, Illinois. Data from the Se-Met crystals were collected at one wavelength (peak). All data were integrated with

DENZO and merged with Scalepack within HKL2000 (*56*). Tables 2.2-2.6 show condensed Scalepack output log files for all four structures, including the Se-Met log file from data used for phasing (Table 2.2). Split into shells of resolution, the log files are used as a tool to understand the quality of the data after being integrated and merged. Specifically, the log files presented in Tables 2.2-2.6 show redundancy and completeness of the shells, and also provide data quality statistics. Understanding these output files is crucial in determining not only inborn problems in the diffraction quality resulting from poor crystals, but also the quality of the choices made during the integration and merging process.

2.2D: Structure Determination

X-rays behave as both a particle and a wave. The ability to "transform" diffraction from the x-ray beam after hitting a protein crystal into an electron density map via the electron density equation (Eq. 1) can only be accomplished when three variables are known.

$$p(xyz) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |F_{hkl}| e^{-2\pi i \left[hx + ky + lz - \varphi_{(hkl)} \right]}$$

Equation 1: Electron Density (ρ) Equation. Red represents the structure factor amplitude. Green represents the indices, hkl, of the unit cell (position). Blue represents the phase term of the electromagnetic wave. V is the unit cell volume.

The three terms shown in color in Equation 1 represent the location of the diffraction spots (known as the hkl's or indices colored in green) in reciprocal space, intensity of the diffraction (known as amplitude colored in red), and finally the phase (φ) of the wave colored in blue. A two-dimensional detector can only acquire measurements for position and amplitude. The inability for the detector to simultaneously measure the phase of the wave is known as the "phase problem".

In order to acquire a three dimensional structure from a two dimensional detector a method for acquiring phases must be addressed. The availability of adjustable wavelengths at synchrotrons has provided a great method for acquiring phases through single wavelength anomalous diffraction (SAD) (*57, 58*). Phasing via SAD methods requires the presence of anomalously diffracting atoms in the structure. This can sometimes be accomplished using the native sulfur atoms (as you will see in chapter 9). However, this is generally more challenging and requires very high resolution (2.0 Å or less) and clean diffraction patterns, including a robust crystal that withstands radiation damage (*59*). Heavy atoms can also be introduced artificially, either through replacement during expression or by heavy atom soaking (*60, 61*). The most common method of heavy atom incorporation is achieved with replacement of methionine residues with selenomethionine (Se-Met). This replacement should not cause any changes to the crystal form (space group and unit cell), and therefore should be "isomorphic" to the native crystal. The Patterson function is used to take advantage of the heavy atom's absorption edge and anomalous diffraction associated with collecting data at that particular wavelength and determines the location of the introduced heavy atoms. Once the locations are determined, phases for the heavy

atom substructure can be determined and the first electron density map can be viewed. From this point, there are many options for model building.

Another option for phasing is called molecular replacement and is now the most common method of phasing (~75%) (12). Molecular replacement is more of a brute force method that rotates and translates a known homologous model (>20% identity) looking for a similar match to the unknown model. Modern search programs, such as Phaser and MOLREP, perform this search by independent rotational and translational searches. First, a proper rotational orientation for the search model is determined in the asymmetric unit by aligning the Patterson maps for the search model with the experimental data. After the correct rotational orientation is determined, the search model undergoes translational movements to better align the Patterson maps. If the resolution is high enough and with today's impressive processing power in computers, in a relatively short time, initial phases can be determined by this rotation and translation search method. Once the phases are refined, an electron density map can be viewed.

The AfNep1 structure was solved using single wavelength anomalous diffraction (SAD) data with the Se-Met protein crystals, collected to 2.0 Å. The AutoSharp program suite was used to obtain the phases using the SAD method, apply solvent flattening and density modification to the initial electron density map, and perform automatic model building with wARP (*62, 63*). The complete model was finalized by rebuilding in COOT and refinement with REFMAC5 (*64, 65*). The model was then used to phase the native data set using MOLREP as part of CCP4i program suite (*66-69*). The ScNep1 structures were solved by molecular replacement in

MOLREP or Phaser using PDB ID 2V3J (*17, 70*). All models were subjected to several rounds of building in COOT and refinement using REFMAC5 or PHENIX (*64, 65, 71, 72*). Translation libration screw-motion (TLS) groups were determined using the TLSMD server (*73*). Refinement statistics are provided in Table 2.7. The coordinates and structure factors have been deposited in the Protein Data Bank (accession codes 3O7B for AfNep1, 3OII, 3OIJ and 3OIN for ScNep1 free dimer, 1RNA complex and 2RNA complex respectively).

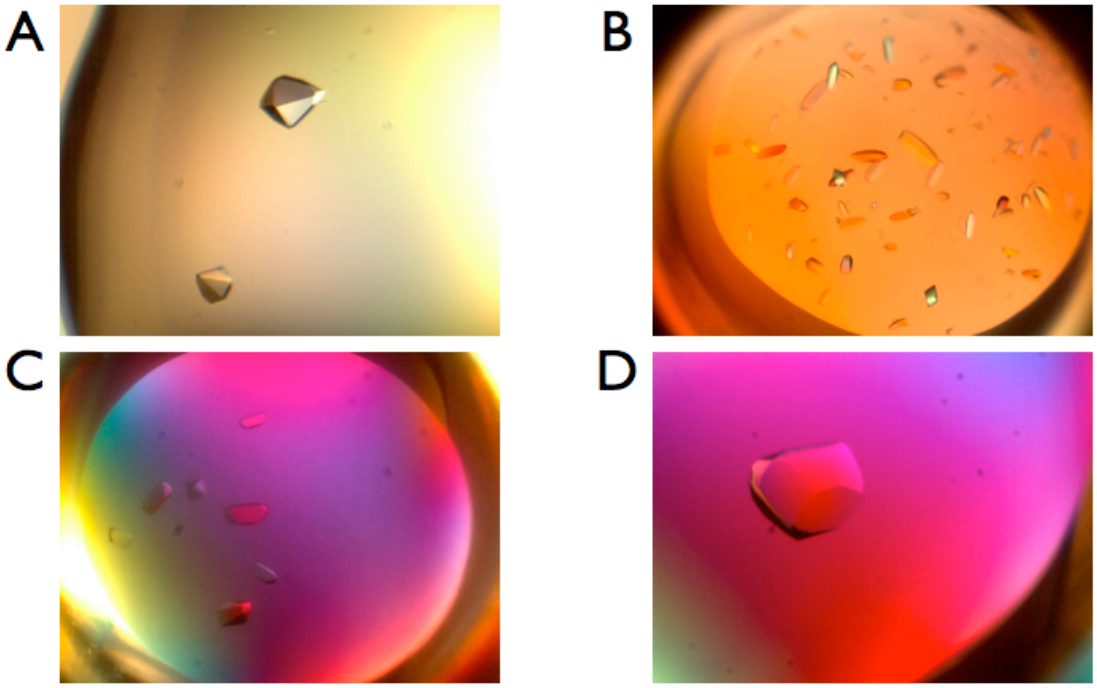


Figure 2.1: AfNep1 and ScNep1 Crystals

A) AfNep1 crystals containing Se-Met, (**B)** ScNep1, (**C**) ScNep1 with one molecule of RNA and (**D)** ScNep1 with two molecules of RNA.

Table 2.1: RNA Oligonucleotides

Green represents consensus yeast consensus sequence (*21*). Red coloring indicates mutations from native yeast 18S RNA.

- **Oligo 1 - 14 base native yeast**

  5'-GCU CUU CAA CGA GG-3'

- **Oligo 2 - mutated yeast *crystallized with Emg1***

  5'-GGG CUU CAA CGC CC-3'

- **Oligo 3 – 12 base native yeast**

  5'-CUC UUC AAC GAG-3'

- **Oligo 4 – 8 base native yeast**

  5'-CUU CAA CG-3'

Table 2.2: *A. fulgidus* Nep1 SeMet Scalepack Logfile

Shell Lower limit	Upper limit	Average Redundancy Per Shell
30.00	4.31	7.3
4.31	3.42	7.5
3.42	2.99	7.6
2.99	2.71	7.6
2.71	2.52	7.6
2.52	2.37	7.6
2.37	2.25	7.6
2.25	2.15	7.6
2.15	2.07	7.6
2.07	2.00	7.6
All	hkl	7.3

Shell		I/Sigma in resolution shells:								
Lower	Upper	% of reflections with I/Sigma less than								
limit	limit	0	1	2	3	5	10	20	>20	total
30.00	4.31	0.0	0.0	0.1	0.1	0.3	0.9	2.7	93.7	96.4
4.31	3.42	0.0	0.1	0.3	0.4	0.6	1.6	4.6	95.2	99.8
3.42	2.99	0.1	0.3	0.6	0.8	1.5	3.3	8.2	91.8	100.0
2.99	2.71	0.4	0.6	1.4	2.0	3.3	6.9	16.1	83.9	100.0
2.71	2.52	0.3	0.8	1.8	2.7	5.0	11.0	23.8	76.2	100.0
2.52	2.37	0.5	1.5	2.6	4.1	7.6	15.7	31.2	68.8	100.0
2.37	2.25	1.0	2.6	4.6	7.1	12.3	22.2	41.4	58.6	100.0
2.25	2.15	1.3	4.0	6.9	10.0	16.3	28.3	49.6	50.3	99.9
2.15	2.07	1.4	4.2	7.4	11.1	16.7	33.0	60.6	39.3	100.0
2.07	2.00	2.4	6.6	11.6	17.0	26.7	45.0	71.6	28.3	99.9
All	hkl	0.7	2.1	3.7	5.5	9.0	16.8	30.9	68.7	99.6

Shell			Average		Norm.	Linear	Square
Lower Limit	Upper Limit	I	error	stat.	Chi**2	R-fac	R-fac
30.00	4.31	2142.9	46.3	16.4	0.593	0.023	0.028
4.31	3.42	2041.8	63.4	14.6	0.586	0.039	0.047
3.42	2.99	938.5	19.7	8.6	0.872	0.037	0.043
2.99	2.71	495.1	11.5	6.3	0.869	0.041	0.042
2.71	2.52	338.8	8.8	5.5	0.918	0.049	0.050
2.52	2.37	253.3	7.4	5.2	0.963	0.058	0.058
2.37	2.25	191.6	6.6	5.1	1.011	0.072	0.069
2.25	2.15	151.6	5.8	5.1	1.150	0.089	0.084
2.15	2.07	116.1	5.6	5.1	1.136	0.110	0.101
2.07	2.00	86.5	5.3	5.0	1.131	0.142	0.127
All	reflections	672.1	18.0	7.7	0.925	0.041	0.040

Table 2.3: *A. fulgidus* Nep1 Native Scalepack Logfile

Shell Lower limit	Upper limit	Average Redundancy Per Shell
30.00	3.12	12.0
3.12	2.48	20.7
2.48	2.17	21.6
2.17	1.97	21.8
1.97	1.83	13.5
1.83	1.72	10.9
1.72	1.63	10.9
1.63	1.56	10.8
1.56	1.50	10.7
1.50	1.45	10.0
All	hkl	14.3

Shell		I/Sigma in resolution shells:								
Lower	Upper	% of reflections with I/Sigma less than								
limit	limit	0	1	2	3	5	10	20	>20	total
30.00	3.12	0.1	0.3	0.4	0.5	0.9	1.8	4.3	94.8	99.2
3.12	2.48	0.1	0.3	0.5	0.6	1.0	2.4	8.2	91.8	100.0
2.48	2.17	0.3	0.6	1.0	1.3	2.1	5.0	12.8	87.2	100.0
2.17	1.97	0.6	1.1	1.6	2.3	3.9	8.2	19.4	80.6	100.0
1.97	1.83	2.3	3.1	4.4	5.6	8.5	17.0	37.9	62.1	100.0
1.83	1.72	3.8	5.8	8.1	10.9	16.3	28.9	55.0	45.0	100.0
1.72	1.63	4.8	7.7	11.6	15.3	22.6	39.6	68.8	31.2	100.0
1.63	1.56	6.0	9.7	14.6	19.9	29.0	49.5	80.0	20.0	100.0
1.56	1.50	7.8	14.1	22.5	30.7	43.4	65.8	89.0	11.0	100.0
1.50	1.45	10.1	20.0	30.3	39.6	55.4	77.5	94.2	5.8	100.0
All	hkl	3.6	6.2	9.4	12.5	18.1	29.3	46.5	53.4	99.9

Shell			Average		Norm.	Linear	Square
Lower Limit	Upper Limit	I	error	stat.	Chi**2	R-fac	R-fac
30.00	3.12	1820.6	49.8	12.0	1.071	0.045	0.053
3.12	2.48	529.1	17.0	2.7	1.156	0.085	0.092
2.48	2.17	272.7	7.9	1.8	1.203	0.096	0.105
2.17	1.97	156.7	4.4	1.7	1.337	0.117	0.123
1.97	1.83	79.7	3.0	1.4	1.191	0.123	0.130
1.83	1.72	41.3	1.9	1.3	1.183	0.139	0.129
1.72	1.63	27.4	1.6	1.3	1.179	0.191	0.176
1.63	1.56	20.2	1.6	1.3	1.163	0.261	0.246
1.56	1.50	13.2	1.5	1.4	1.088	0.384	0.352
1.50	1.45	9.3	1.6	1.4	1.085	0.517	0.450
All	reflections	305.8	9.3	2.7	1.182	0.075	0.063

Table 2.4: *S. cerevisiae* Nep1 Scalepack Logfile

Shell Lower limit	Upper limit	Average Redundancy Per Shell
30.00	3.88	7.6
3.88	3.08	7.9
3.08	2.69	8.0
2.69	2.44	8.0
2.44	2.27	7.9
2.27	2.13	7.6
2.13	2.03	6.9
2.03	1.94	5.6
1.94	1.86	4.3
1.86	1.80	3.2
All	hkl	6.9

Shell		I/Sigma in resolution shells:								
Lower	Upper	% of reflections with I/Sigma less than								
limit	limit	0	1	2	3	5	10	20	>20	total
30.00	3.88	0.4	1.0	1.5	1.9	2.7	4.7	12.2	87.3	99.5
3.88	3.08	0.3	1.0	2.0	2.9	4.6	9.6	25.4	74.6	100.0
3.08	2.69	1.6	3.2	5.4	7.5	11.7	21.5	49.0	50.9	99.9
2.69	2.44	2.8	5.9	9.3	12.6	19.0	35.1	64.2	35.7	99.9
2.44	2.27	3.3	7.2	12.2	17.2	26.4	45.5	75.2	24.3	99.5
2.27	2.13	3.2	9.5	16.8	23.6	35.7	58.9	85.5	13.8	99.3
2.13	2.03	4.1	13.6	24.7	36.1	51.6	76.2	95.9	3.2	99.1
2.03	1.94	8.0	22.4	37.4	49.6	65.2	85.0	95.5	2.2	97.7
1.94	1.86	11.1	29.5	45.6	56.6	68.4	79.7	84.2	0.6	84.9
1.86	1.80	9.6	28.5	41.9	49.3	56.8	62.1	63.1	0.0	63.1
All	hkl	4.4	12.0	19.4	25.4	33.8	47.3	64.5	29.9	94.4

Shell			Average		Norm.	Linear	Square
Lower Limit	Upper Limit	I	error	stat.	Chi**2	R-fac	R-fac
30.00	3.88	768.3	23.5	9.2	0.709	0.040	0.047
3.88	3.08	411.5	14.3	6.2	0.657	0.052	0.056
3.08	2.69	138.4	6.0	3.9	0.579	0.069	0.070
2.69	2.44	77.9	4.0	3.3	0.549	0.088	0.084
2.44	2.27	55.7	3.7	3.4	0.577	0.121	0.117
2.27	2.13	39.4	3.6	3.5	0.581	0.170	0.167
2.13	2.03	28.0	4.0	3.7	0.519	0.218	0.204
2.03	1.94	18.1	4.0	3.9	0.523	0.279	0.251
1.94	1.86	12.1	4.4	4.3	0.530	0.350	0.302
1.86	1.80	7.9	4.7	4.6	0.441	0.403	0.369
All	reflections	169.4	7.5	4.6	0.585	0.061	0.052

Table 2.5: *S. cerevisiae* Nep1 with One Molecule of RNA Scalepack Logfile

Shell Lower limit	Upper limit	Average Redundancy Per Shell
30.00	4.09	7.5
4.09	3.25	7.9
3.25	2.84	8.0
2.84	2.58	8.1
2.58	2.39	8.1
2.39	2.25	8.0
2.25	2.14	7.5
2.14	2.05	6.2
2.05	1.97	5.1
1.97	1.90	4.2
All	hkl	7.1

Shell		I/Sigma in resolution shells:								
Lower	Upper	% of reflections with I/Sigma less than								
limit	limit	0	1	2	3	5	10	20	>20	total
30.00	4.09	0.5	0.7	0.8	0.9	1.1	2.4	10.9	88.1	99.0
4.09	3.25	0.7	1.1	1.5	1.9	3.3	7.2	23.4	76.5	100.0
3.25	2.84	2.1	3.1	4.7	6.6	10.3	20.0	38.9	61.1	100.0
2.84	2.58	3.6	6.4	9.7	13.1	20.2	35.7	56.9	43.1	100.0
2.58	2.39	5.1	9.2	15.0	20.7	29.7	47.8	72.6	27.4	100.0
2.39	2.25	5.8	11.9	20.0	27.0	38.5	59.1	82.0	18.0	100.0
2.25	2.14	7.7	17.5	27.9	37.0	50.8	71.2	89.5	10.5	100.0
2.14	2.05	9.4	23.0	37.1	47.3	61.2	78.7	92.7	5.0	97.7
2.05	1.97	10.4	27.7	42.2	50.7	64.1	79.3	88.1	2.5	90.6
1.97	1.90	12.7	30.3	44.0	53.8	64.4	74.7	79.5	0.7	80.2
All	hkl	5.7	12.9	20.0	25.6	33.9	47.1	62.9	33.9	96.8

Shell			Average		Norm.	Linear	Square
Lower Limit	Upper Limit	I	error	stat.	Chi**2	R-fac	R-fac
30.00	4.09	761.0	22.7	7.4	1.083	0.047	0.052
4.09	3.25	365.5	13.4	4.4	1.040	0.062	0.065
3.25	2.84	116.8	3.7	2.7	1.266	0.077	0.074
2.84	2.58	52.2	2.3	2.2	1.103	0.105	0.097
2.58	2.39	32.4	2.1	2.1	0.911	0.149	0.137
2.39	2.25	23.0	2.1	2.1	0.787	0.194	0.178
2.25	2.14	16.0	2.2	2.1	0.723	0.256	0.232
2.14	2.05	10.9	2.4	2.4	0.682	0.301	0.285
2.05	1.97	7.7	2.4	2.3	0.648	0.336	0.313
1.97	1.90	5.1	2.2	2.1	0.613	0.392	0.374
All	reflections	148.2	5.8	3.0	0.928	0.066	0.057

Table 2.6: *S. cerevisiae* Nep1 with Two Molecules of RNA Scalepack Logfile

Shell Lower limit	Upper limit	Average Redundancy Per Shell
30.00	6.23	6.7
6.23	4.95	7.0
4.95	4.33	7.2
4.33	3.93	7.2
3.93	3.65	7.3
3.65	3.44	7.2
3.44	3.27	6.9
3.27	3.12	5.7
3.12	3.00	4.1
3.00	2.90	2.9
All	hkl	6.3

Shell		I/Sigma in resolution shells:								
Lower	Upper	% of reflections with I/Sigma less than								
limit	limit	0	1	2	3	5	10	20	>20	total
30.00	6.23	0.9	1.7	2.5	3.4	5.3	10.4	21.7	78.0	99.8
6.23	4.95	0.6	2.9	4.9	6.6	9.4	16.8	35.9	64.1	100.0
4.95	4.33	1.0	3.2	4.9	6.0	9.3	17.2	37.5	62.4	99.9
4.33	3.93	1.0	3.2	6.3	8.3	13.5	25.6	48.9	51.1	100.0
3.93	3.65	1.4	5.2	8.9	12.0	19.5	34.7	59.8	40.2	100.0
3.65	3.44	2.1	7.3	14.0	19.8	29.3	50.3	77.9	22.1	100.0
3.44	3.27	3.8	11.3	20.9	29.3	43.8	68.3	90.7	9.2	99.9
3.27	3.12	5.8	17.6	32.3	43.4	61.7	84.4	96.6	2.1	98.7
3.12	3.00	8.1	24.6	40.8	54.0	72.5	88.8	94.0	0.6	94.6
3.00	2.90	8.2	30.1	47.1	56.6	67.4	75.0	76.5	0.1	76.5
All	hkl	3.2	10.6	18.0	23.6	32.7	46.6	63.4	33.6	97.0

Shell			Average		Norm.	Linear	Square
Lower Limit	Upper Limit	I	error	stat.	Chi**2	R-fac	R-fac
30.00	6.23	145.2	3.8	2.0	1.010	0.049	0.064
6.23	4.95	52.5	1.7	1.3	0.856	0.061	0.067
4.95	4.33	58.4	2.0	1.5	1.038	0.072	0.079
4.33	3.93	41.8	1.6	1.5	0.961	0.080	0.082
3.93	3.65	32.1	1.5	1.5	0.910	0.097	0.095
3.65	3.44	20.9	1.5	1.5	0.743	0.133	0.126
3.44	3.27	12.7	1.5	1.5	0.641	0.189	0.178
3.27	3.12	7.9	1.6	1.6	0.568	0.263	0.247
3.12	3.00	5.6	1.8	1.8	0.522	0.293	0.279
3.00	2.90	3.9	2.0	2.0	0.492	0.361	0.336
All	reflections	40.3	1.9	1.6	0.824	0.078	0.070

Table 2.7: Nep1/Emg1 data collection and refinement statistics

	afNep1 Se-Met SAD	**afNep1**	**scNep1**	**scNep1 + 1RNA**	**scNep1 + 2RNA**
PDB code		**3O7B**	**3OII**	**3OIN**	**3OIJ**
Data collection					
Space group	$P3_12_1$	$P3_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions					
a, *b*, *c* (Å)	70.3, 70.3, 95.3	70.1, 70.1, 95.2	46.0, 84.7, 112.5	44.8, 88.7, 115.8	83.4, 91.9, 96.9
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 90.0	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Molecules/Asym. unit	1 (1 Protein)	1 (1 Protein)	2 (2 Protein)	3 (2 Protein: 1 RNA)	4 (2 Protein: 2 RNA)
Wavelength (Å)	0.97919	0.99998	1.00931	1.00931	1.00931
Resolution (Å)	30-2.0	30-1.45	30-1.8	30-1.9	30-2.9
R_{sym} (last shell)	0.041 (0.142)	0.075 (0.517)	0.061 (0.403)	0.066 (0.392)	0.078 (0.361)
$I/\sigma I$	37.3 (16.3)	32.9 (5.8)	22.6 (1.7)	25.5 (2.3)	21.2 (2.0)
Completeness (%)	99.6 (99.9)	99.0 (100.0)	94.4 (63.1)	96.8 (80.2)	97.0 (76.5)
Redundancy	7.6 (7.6)	14.3 (10.0)	6.9 (3.2)	7.1 (4.2)	6.3 (2.9)
Refinement					
Resolution (Å)		29-1.45	20-1.85	28-1.9	30-3.0
R_{work}/R_{free} (%)		19.8/21.1	17.1/21.6	19.2/23.0	17.4/25.2
# protein residues monomers A/B		216	216/208	217/217	211/217
No. non protein molecules					
Nucleic Acid (bases)		0	0	14	28
Ligand (SAH)		1	2	1	2
Solvent		242	224	258	64
Mean B-factors $(Å)^2$ (TLS refined)		23.3	35.0	48.1	68.9
RMS deviations					
Bond lengths (Å)		0.010	0.021	0.008	0.008
Bond angles (°)		1.281	1.881	1.069	1.257
Ramachandran plot					
Most favorable		94.2%	93.4%	92.8%	87.5%
Additionally allowed		5.8%	6.3%	7.0%	12.5%
Generously allowed		0.0%	0.3%	0.3%	0.3%
Disallowed		0.0%	0.0%	0.0%	0.0%

Chapter 3: Description of AfNep1 and ScNep1 in the Free State

3.1: Overview

Structural analysis of AfNep1 and ScNep1 supported previously reported structures that Nep1 does indeed belong to the SPOUT family of methyltransferases and contained all of the characteristic family traits, including the knotted methyltransferase fold (*16, 17*). This chapter will present a structural overview of Nep1, highlighting Nep1's relationship to the SPOUT family of methyltransferases. Secondly, all SPOUT family proteins have been shown to be dimers (*19*), as Nep1 is not an exception, a description of the dimer interface combined with comparisons between AfNep1 and ScNep1 will be presented in detail. Research on the characterization of the Nep1 dimer interface was progressing very quickly and was near write up and submission, however, the competition was fierce and the direction of the project had to adapt (Chapter 4) after other laboratories released initial structural descriptions and also supported the presence of an obligate dimer (*16, 17*).

3.2: Basic Structural Description

The overall structure of the Nep1 protein, shown in Figure 3.1, consists of a central β-sheet surrounded by α-helices on either side, characteristic of a Rossmann fold (*16, 17*). The central β-sheet is comprised of six β-strands, five of which are parallel to each other and one anti-parallel (β3), with an additional two loosely ordered strands completing the sheet (β4, β5). The order in which sequential β-strands form the β-sheet is β5-β4-β3-β2-β1-β7-β6-β8. The helices within the

Rossmann domain are located between β1 and β2 (αA, αAb, αd´), β3 and β4 (αc, Hc) and after β8 (αE). Two additional helices, Hb´ and αa´ are present in a small Nep1 specific 40 amino-acid insertion between β1 and α2 (residues 48–88). This insertion and the orientation of αa´ is characteristic of Nep1 SPOUT methyltransferases, and also exhibits the most difference between the archaeal and yeast structures. This domain is moderately conserved based on sequence comparison with eukaryotic Nep1 sequences (Fig. 3.1). Electron density was observed for all the residues except 1–27 for both monomers, 56–64 for monomer A and 54–65 and 153–157 for monomer B, of ScNep1, and for all but the last three for AfNep1. In both models, density was observed in the SAM-binding pocket into which SAH was modeled.

3.3: Dimer Formation

Although the structure of Nep1 from yeast has been previously reported, the deposited structure (PDB ID 2V3J) was monomeric (*17*). In the crystal structure of ScNep1 bound to SAH presented here, the molecule is dimeric. In both archaea and yeast, a large interacting surface is observed between Nep1 molecules that buries surface area in excess of 845 Å^2 for AfNep1 and around 1475 Å^2 for ScNep1. The dimer interface is formed primarily between residues on helices αA and αE and the N-terminal end of αc'. The interactions form in a 2-fold symmetric interface that can be seen in Figure 3.2A and Figure 3.2B (Fig. 3.3 for AfNep1). The residues at the dimer interface of ScNep1 are highly conserved among eukaryotes, but less so among archaeal organisms (Figs. 3.4 and 3.5). The ScNep1 interface consists of a mixture of electrostatic interactions and hydrogen bonds on the edges and a few

hydrophobic interactions at the center, and several water molecules were observed in the charged regions of the interface. This shows a less tight dimer interface than that observed for the AfNep1 interface, which contains a stabilizing disulfide bond, commonly used to stabilize protein structure in thermophiles (Fig. 3.3) (*74*). Interestingly, in ScNep1, the side chain of the C-terminal Leu 252 forms hydrophobic packing interactions at the interface and the carboxyl group at the end of the chain forms a hydrogen bond with the side chain of Ser 228 and Asn 229 and the backbone amide nitrogen of Asn 229 of the other monomer (Fig. 3.2B and 3.2C). Therefore, the C-terminus participates in dimer formation. In the previously reported structure of ScNep1, the construct used contained a C-terminal histidine tag that may have disrupted this interaction and weakened the dimer to some extent, resulting in monomers in the crystals (*17*).

3.4: Chapter Summary

The conserved dimer interface between the monomers of both AfNep1 and ScNep1 is seen throughout all members of the SPOUT family (*19*). While we were working on exploring this dimer interface, other research laboratories published work confirming Nep1's conserved dimer formation. The focus of research was no longer to determine a novel structure, but to determine Nep1's involvement with target 18S RNA. Focus shifted to crystallization of Nep1 in complex with RNA.

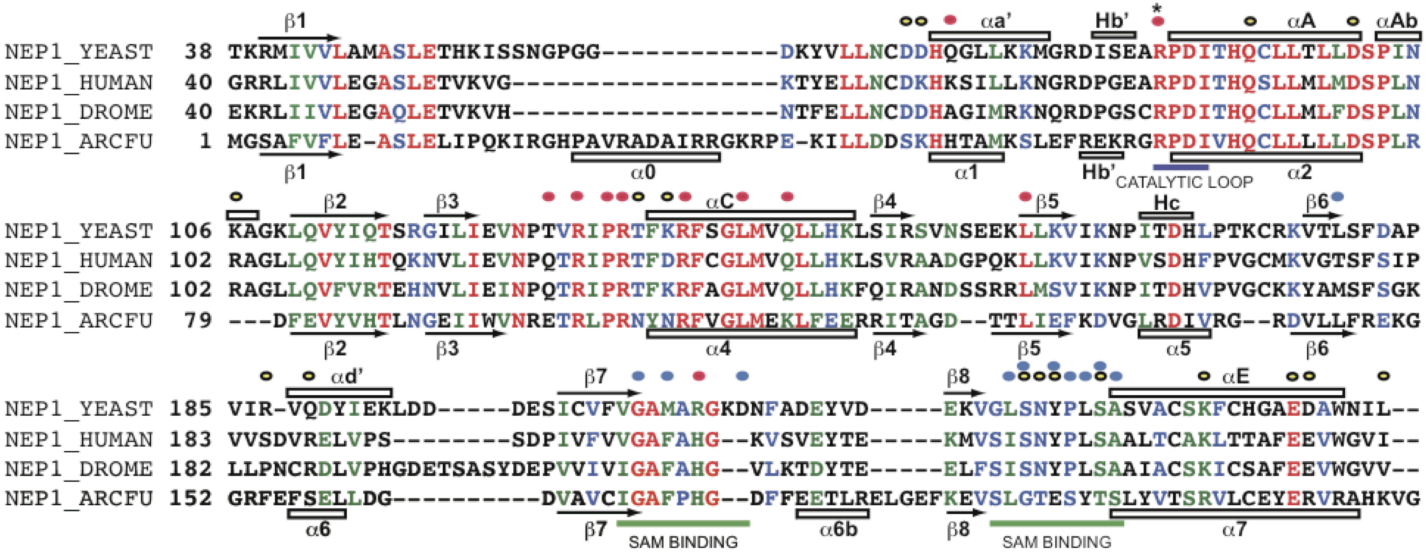

β1
αa'
Hb'
*
αA
αAb
NEP1_YEAST 38 TKRMIVVLAMASLETHKISSNGPGG----------DKYVLLNCDDHQGLLKKMGRDISEARPDITHQCLLTLLDSPIN
NEP1_HUMAN 40 GRRLIVVLEGASLETVKVG----------KTYELLNCDKHKSILLKNGRDPGEARPDITHQSLLMLMDSPLN
NEP1_DROME 40 EKRLIIVLEGAQLETVKVH----------NTFELLNCDDHAGIMRKNQRDPGSCRPDITHQCLLMLFDSPLN
NEP1_ARCFU 1 MGSAFVFLE-ASLELIPQKIRGHPAVRADAIRRGKRPE-KILLDDSKHHTAMKSLEFREKRGRPDIVHQCLLLLDSPLR
β1
α0
α1
Hb'
CATALYTIC LOOP
α2
β2
β3
αC
β4
β5
Hc
β6
NEP1_YEAST 106 KAGKLQVYIQTSRGILIEVNPTVRIPRTFKRFSGLMVQLLHKLSIRSVNSEEKLLKVIKNPITDHLPTKCRKVTLSFDAP
NEP1_HUMAN 102 RAGLLQVYIHTQKNVLIEVNPQTRIPRTFDRFCGLMVQLLHKLSVRAADGPQKLLKVIKNPVSDHFPVGCMKVGTSFSIP
NEP1_DROME 102 RAGLLQVFVRTEHNVLIEINPQTRIPRTFKRFAGLMVQLLHKFQIRANDSSRRLMSVIKNPITDHVPVGCKKYAMSFSGK
NEP1_ARCFU 79 ---DFEVYVHTLNGEIIWVNRETRLPRNYNRFVGLMEKLFEERRITAGD--TTLIEFKDVGLRDIVRG--RDVLLFREKG
β2
β3
α4
β4
β5
α5
β6
αd'
β7
β8
αE
NEP1_YEAST 185 VIR-VQDYIEKLDD-----DESICVFVGAMARGKDNFADEYVD-----EKVGLSNYPLSASVACSKFCHGAEDAWNIL--
NEP1_HUMAN 183 VVSDVRELVPS--------SDPIVFVVGAFAHG--KVSVEYTE-----KMVSISNYPLSAALTCAKLTTAFEEVWGVI--
NEP1_DROME 182 LLPNCRDLVPHGDETSASYDEPVVIVIGAFAHG--VLKTDYTE-----ELFSISNYPLSAAIACSKICSAFEEVWGVV--
NEP1_ARCFU 152 GRFEFSELLDG----------DVAVCIGAFPHG--DFFEETLRELGEFKEVSLGTESYTSLYVTSRVLCEYERVRAHKVG
α6
β7
SAM BINDING
α6b
β8
SAM BINDING
α7

Figure 3.1: Nep1 Sequence Alignment

Sequence alignment from of Nep1 from *S. cerevisiae* (NEP1_YEAST), Homo sapiens (NEP1_HUMAN), *Drosophila melanogaster* (NEP1_DROME) and *A. fulgidus* (NEP1_ARCFU). Yellow, red and blue dots indicate residues involved in dimerization, RNA binding and SAM binding respectively. Secondary structure for yeast (top) and *A. fulgidus* (bottom) shown as open bars for α-helices and arrows for β-strands and numbered according to Leulliot *et al*. for yeast, and Taylor *et al*. for *A. fulgidus* (*16, 17*).

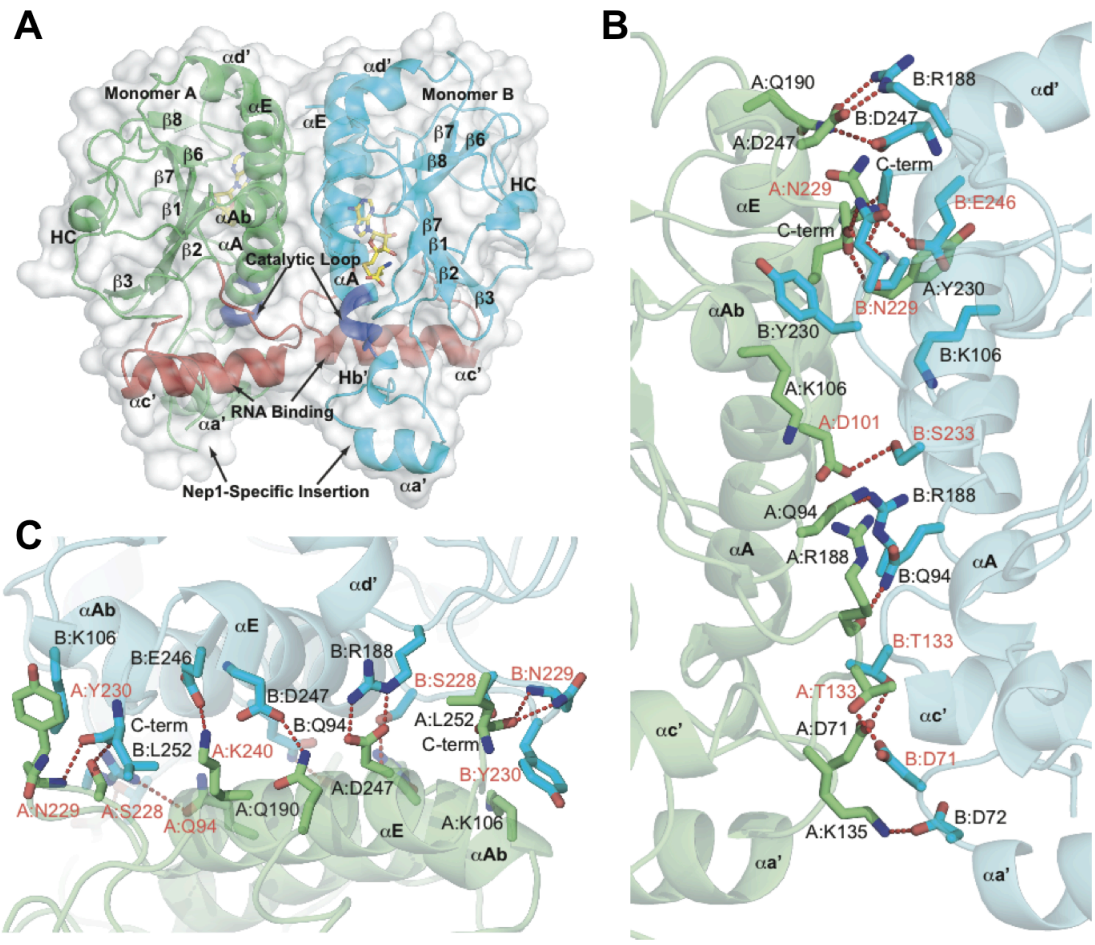


Figure 3.2: Overall Structure of ScNep1 and Dimer Interface

A) Overall view of the ScNep1 dimer. RNA binding and catalytic regions of the molecules are colored red and dark blue respectively. The molecules of SAH are shown in sticks. Secondary structure elements are labeled according to Leulliot et al. (*17*). **B**) Front view of the dimerization interface of ScNep1. Residues are labeled with chain names and residue numbers. **C)** Top view of the interface showing the interaction of the C-terminus in the interface. Residues that are identical in eukaryotic sequences are labeled in red text. All 3D structure figures were generated in PyMOL (www.pymol.org) (*75*).

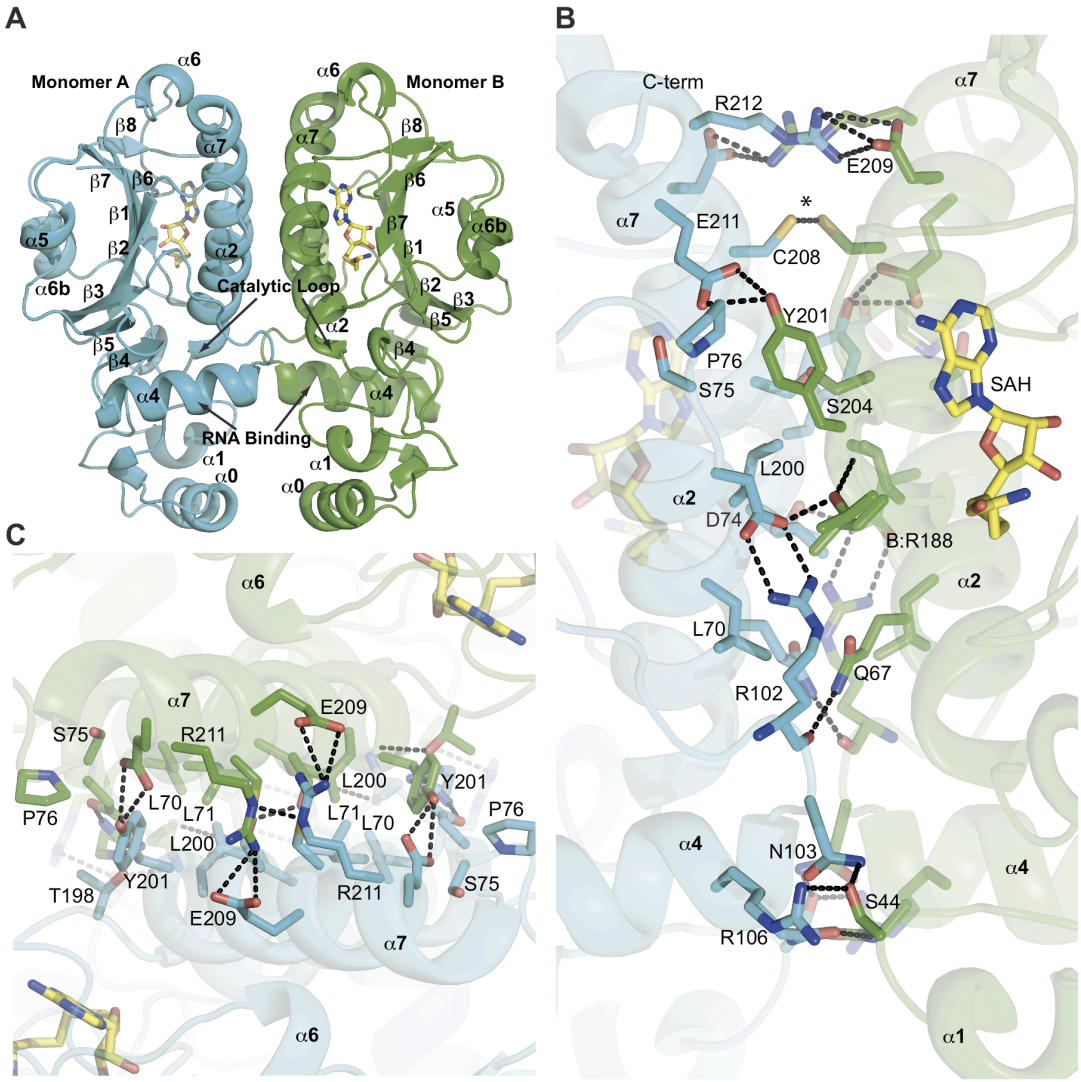


Figure 3.3: The AfNep1 Dimer

A) The symmetric AfNep1 dimer generated by symmetry operation of the asymmetric unit. Bound SAH molecules (not added to crystallization conditions) shown in yellow sticks. **B)** Close-up of dimerization interface. Cysteine involved in dissulfide bond labelled with an asterisk. **C)** Top view of the dimerization interface.

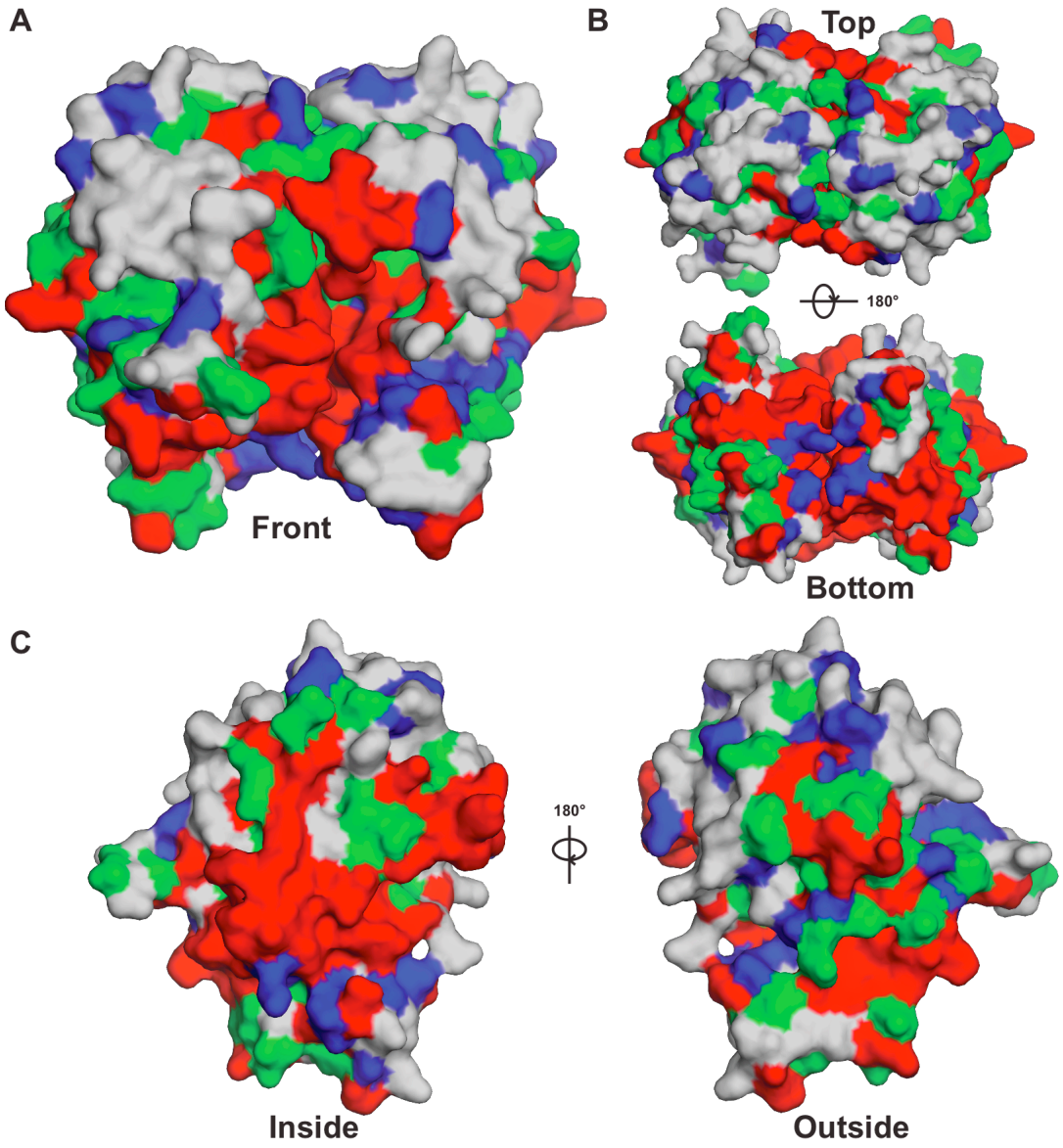


Figure 3.4: Conservation Maps on the Surface of ScNep1

Residues are colored based on conservation in alignments with human and fruit fly Nep1 sequences using ClustalW (red = identical; green = highly conserved; blue = weakly conserved). The "front" view in **A**. is rotated 90° in either direction around the x-axis to produce the orientations in **B**. In **C**, "inside" refers to the dimerization interface, and "outside" is produced by a 180° rotation around the y-axis.

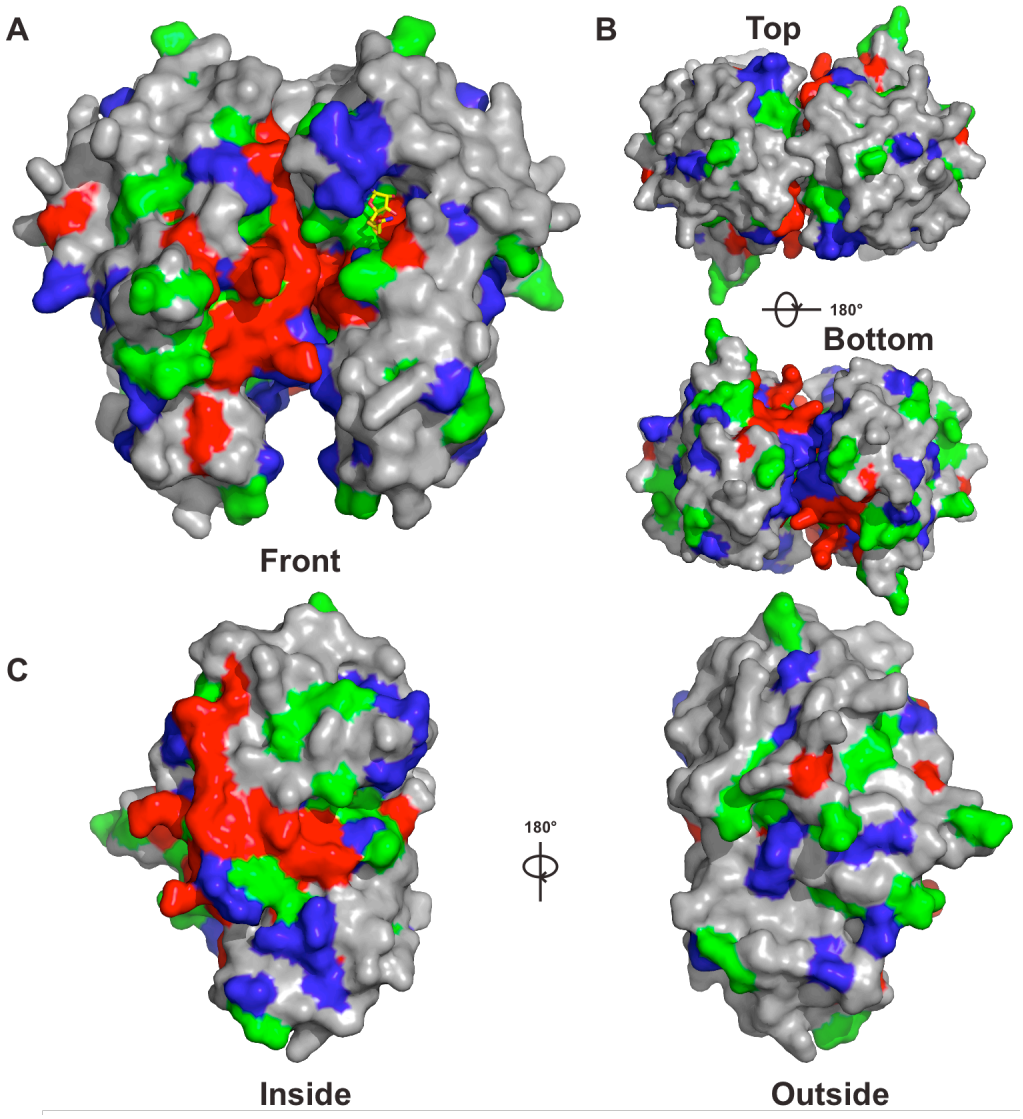


Figure 3.5: Conservation Maps on the Surface of AfNep1

A-C) Residues are colored based on conservation in alignments with eukaryotic Nep1 sequences using ClustalW (red = identical; green = highly conserved; blue = weakly conserved) (*46*).

Chapter 4: Nep1 in Complex with RNA

4.1: Overview

Our structures of ScNep1 are the first examples of a SPOUT methyltransferase bound to its target RNA. Even more interesting is Nep1 is the only member of the SPOUT family known to date that methylates pseudouridine (*22*). Presented in this chapter is a base-by-base description of the interactions between ScNep1 and the 14 base RNA fragment. Conformational changes between free ScNep1 versus ScNep1 bound to RNA and comparisons between ScNep1 with one molecule of RNA versus ScNep1 with two molecules of RNA are also presented in dedicated sections within this chapter.

4.2: Interaction of ScNep1 with RNA

Using the binding sequences for ScNep1 obtained through RNA three-hybrid screening by Buchhaupt et al. (*21*), a series of oligonucleotides were designed and tested for binding to ScNep1 in stoichiometric complexes. It was determined that an ideal complex was obtained when ScNep1 and a 14 base RNA fragment with the sequence $G_1G_2G_3C_4U_5U_6C_7A_8A_9C_{10}G_{11}C_{12}C_{13}C_{14}$ were mixed in a 2:2 (monomer:RNA) ratio. Crystals were obtained for Nep1 bound to this 14 base RNA fragment in two crystal forms (Fig. 2.1). In one crystal form, a dimer of ScNep1 is bound to one RNA molecule (Fig. 2.1A). In the second crystal form, the dimer binds two separate RNA molecules (Fig. 2.1B). In both cases, the RNA forms a stem loop structure in which eight bases forming canonical base pairs in the stem and six bases

in the loop (UUCAAC) have base specific contacts with the protein. Electron density was not observed for residues 1-27 and 56-64 of both protein monomers in the RNA complexes, and residues 76-82 of monomer A of the two-RNA complex. There are no significant differences between the RNA molecules bound in the two complexes. In the comparison of the sequences of RNA found to bind Nep1 by RNA three-hybrid screens, all the sequences contained the consensus sequence C/UUCAAC, and the most tightly binding were those sequences that contained the consensus sequence within the loop of a predicted stem-loop (*21*). These six bases form the basis for specific recognition by the protein. In the crystal containing the one RNA per dimer complex, the empty RNA-binding site is involved in four hydrogen bonds with the RNA backbone from a symmetry-related complex.

In both complexes, the RNA binds on highly positively charged surfaces of the dimer as predicted in previous reports (*16, 17*). However, the protein contains significant positively charged surface that is not contacted by RNA in these structures (Fig. 4.2). This positively charged surface is conserved in AfNep1 as well. This is consistent with the idea that Nep1 binds to RNA on the pre-ribosome, which makes it likely that it will contact significantly more RNA than is observed with these minimal constructs. The regions that are not contacted by RNA in these structures contain many conserved residues as well, indicating that this surface is important for Nep1 function (Fig. 3.3).

ScNep1 provides several key interactions that confer the observed specificity for a UUCAAC sequence (Fig. 4.3, schematic in Fig. 4.4). The first base of the consensus sequence, U_5, is contacted through base-specific contacts through bidentate

hydrogen bond interactions between the NH1 and NH2 of conserved Arg 136 of monomer A (A:Arg 136 NH1 and NH2) and the C2 carbonyl oxygen (O2) of the uridine base (Fig. 4.3A, 4.4). There are no contacts between the protein and C4 carbonyl oxygen (O4), which explains the observation that a cytosine, which differs from uracil solely by an amine group at C4, can substitute at this position (*21*).

The next uracil in the consensus sequence, U_6, is the site of modification by the enzyme. It is flipped out from the loop and bound in a pocket between the two monomers, near the bound SAH molecule, in the enzyme active site (Fig. 4.3A, B). It is recognized by a direct hydrogen bond with A:Arg 132 NE (Nε of arginine 132 of monomer A) at the U_6 O2 (C2 carbonyl oxygen of U_6), and water-mediated hydrogen bond interactions between the U_6 O4 and the backbone amide of B: Ser 233 and the U_6 O4 and N3 and OE (side chain carbonyl oxygen) of conserved B: Gln 94. Arg 132 is held firmly in position by hydrogen bonding between Arg 132 NH1 and NH2 and conserved A: Asp 101 OD1 and OD2. In addition, hydrogen bonds are observed between the U_6 O2 and NH1 and NH2 of B:Arg 88 and between the ribose O4 and the NH1 of B:Arg 88. This would hold U_6 firmly in position, and may serve as a means to stabilize negative charge on the base during methyl transfer. Arg 88, in turn, is held in place by interactions between its NH2 and NE and the OD1 and OD2 of conserved A:Asp 90. Asp 90 is the equivalent of the human Asp 96 that is mutated to glycine in Bowen-Conradi syndrome. That mutation would have the effect of destabilizing the interaction with the U_6 substrate both through disruption of the interaction with the Arg 88 and hydrophobic contacts with Ile 91 through repositioning. Finally C5 of the uridine base is positioned a mere 3.9 Å away from the sulfur atom of SAH, in perfect

position for a methyl transfer to the C5 position of the uracil (Fig. 4.3B). If a pseudouridine were in its place, then N1, which is an equivalent position to C5, would be the target of methylation.

In the remaining consensus sequence, C_7 is recognized through backbone interactions between the carbonyl oxygen, O2, and the backbone amides on either side of A: Arg 132, and a base stacking interaction with A: Arg 136. A_8 is sandwiched between a stacking interaction with conserved A: Arg 129 and van der Waal's contact with conserved A: Ile 159, and donates a hydrogen bond from N6 to the backbone carbonyl oxygen of A: Thr 127. A_8 N6 is also involved in a water-mediated hydrogen bond interaction with the side chain OH of Thr 127 and A_8 N1 is in hydrogen bonding distance of the backbone amide of A: Arg 129. A_9 is recognized through hydrogen bonds between the backbone carbonyl of A: Arg 129 and the N6, a van der Waal's interaction between the side chain of A: Val 128 and the C2, and a stacking hydrophobic interaction with conserved A: Leu 140. Although Val 128 is not highly conserved, it is Val, Thr or Cys in other sequences, which would be able to maintain the interaction with A_9 C2. Finally C_{10} is specified by two hydrogen bonds with conserved A: Arg 136, one between the Arg NH2 and the C_{10} N3, and the other between the Arg NE and C_{10} O2. There is also an interaction between the C_{10} ribose O2' and the backbone carbonyl of A: Arg136. Arg 136 contacts C_{10} via three hydrogen bonds, U_5 via two hydrogen bonds, and stacks in between A_9 and C_7. Consequently, Arg 136 is invariant in all sequences. In a functional analysis by Taylor et al., mutations of Arg 136, as well as Arg 88, Arg 129, and Arg 132 to Ala

all result in diminished RNA binding activity as measured by yeast three hybrid assays with consensus sequences (16).

Of the eight bases that form the stem in canonical Watson-Crick base pairs, only one, G_{11}, is contacted at the base by a protein residue. B: Glu 74 accepts a hydrogen bond from G_{11} N2, and contacts the G_{11} ribose O2' via a water-mediated interaction (Fig. 4.5). Other backbone interactions between the protein and the RNA include contacts between A: Gln 143 and C_{10} O2' and phosphate between C_{10} and A_9, A: Arg 129 and C_{11} O2P, A: Arg 211 and the phosphate between U_5 and U_6. Of these, Arg 129 and Gln 143 are highly conserved, while Arg 211 is only weakly conserved (Fig. 3.1A). Based on this data, we can conclude that very little specificity is determined in the stem portion of the RNA.

4.3: Catalytic Residues in the Active Site

The methylation reaction that Nep1 catalyzes would require the deprotonation of N1 of the pseudouridine so that it could attack the methyl group from SAM. Stabilization of a negative charge on the uridine ring would result in a decrease of the pK_a, which was measured at 8.97 for free pseudouridine (*76*). The likely candidate for this stabilization would be invariant Arg 88, for which both primary amine nitrogens are within hydrogen bonding distance of O2, the C2 carbonyl oxygen (3.3 and 2.8 Å); (Fig. 4.3B), and which is stabilized by bidentate hydrogen bond interactions with invariant Asp 90, which is mutated in Bowen-Conradi Syndrome (*77*). Mutation of Arg 88 to Asp resulted in a significant loss of Nep1 RNA binding activity (*17*). Arg 88 and Asp 90 are located at the beginning of αA, and are surrounded by Pro 89 and Ile 91, which are also invariant. Ile 91 is in van der Waals contact distance of the

uridine ring in the active site. This region is labeled as the "Catalytic Loop" in Fig. 3.1A and 3.1B. Arg 132 may also provide further stabilization, since it is also within hydrogen bonding distance (3.0 Å) of O2. These interactions would increase deprotonation at the N1 through localization of the negative charge to O2 where it would be stabilized by interaction or proton donation by Arg 88. This would in turn enable N1 to accept the methyl group from SAM.

4.4: Free Nep1 vs Nep1 Bound to RNA

When a monomer of free Nep1 is compared to monomers from the RNA Nep1 complex, an RMSD of 0.76 Å is observed for 415 out of 426 residues. Most of the interaction between the RNA and the protein are mediated by the helix αc' and adjacent loops on the monomer that does not provide the SAM to the active site. However, the most significant conformational changes are seen in the small Nep1 specific inserted domain on the SAM binding side of the dimer interface (Fig. 4.5A, B). This domain, consisting of a 3-10 helix (Hb'), small α-helix (αa'), and a β-hairpin, undergoes a shift in positioning towards the dimer interface when RNA is bound. In addition, there is alteration of overall secondary structure, as the structure changes to accommodate interactions with the RNA (Fig 4.5A, B). The domain, which is packed against α4, houses Arg 88, which may be important for catalysis, and Glu 74 that interacts with G11. The buried surface area in the RNA bound structure (1338 $Å^2$) is slightly decreased, due to an opening of the interface to accommodate the RNA, which buries an additional 721 $Å^2$ on the monomer containing the methyl donor and 351 $Å^2$ on the other monomer.

4.5: Two RNA vs One RNA

Both the 1:1 and 1:2 (ScNep1 dimer:RNA) complexes for the complexes between Nep1 and RNA were obtained using the same preparation of Nep1 to RNA in a 2:2 (ScNep1 monomer:RNA) ratio. The isolation of these two complexes in the crystal structures suggests that both complexes exist in solution, which implies differences in the binding affinities between the two sites. Comparisons of the structures of these complexes reveal differences in the structures of the RNA binding sites as a function of occupancy with RNA (Fig. 4.5C, D). The RMSD for the alignment of the protein molecules in the complexes averages 0.85 Å for 421 residues of the protein chains, with the largest difference being between chain A of the 1:1 complex and chain B of the 2:1 complex. This number is larger than the RMSD for the comparison between the Nep1/1RNA complex and the free enzyme, indicating a larger conformational difference between the Nep1/1RNA and Nep1/2RNA complexes than between the free enzyme and the Nep1/1RNA complex. Upon closer inspection, we observed that the positioning of one monomer relative to the other is slightly different between the two complexes. In the Nep1/2RNA complex, the two monomers are slightly closer together near the RNA binding site, indicating that RNA binding pulls the two monomers closer together. As a result, the SAH molecule is pulled closer to the RNA. The SAH molecule is contacted mostly by backbone interactions, but one residue, Asp 214, hydrogen bonds the ribosyl 3'OH via one of its side-chain carbonyl oxygens (Fig. 4.6). In the Nep1/2RNA complex, for one monomer site the Asp 214 side-chain is rotated away from the SAM binding site and no longer contacts the bound SAM (Fig. 4.7A, B). This results in a slightly altered

positioning of the methionyl moiety in the active site, with the sulfur atom positioned farther away from the uridine base (Fig. 4.7A, B). In addition, the SAM binding pocket on the side of the dimer not bound to RNA in the Nep1/1RNA complex is not visibly occupied by SAH. Based on these observations, and the presence of additional conserved positive charged surface (Fig. 4.2), it is possible that both RNA binding sites would be occupied on rRNA, perhaps with varied binding affinity or specificity, and additional binding surface would be used to interact on the surface of the rRNA.

4.6: Binding Assays with Nep1 and RNA

Many attempts were made to determine the binding affinity of Nep1 to the target sequence. Electrophoretic mobility gel shift assay (EMSA) and anisotropy are described. For gel shift, a 12 base RNA oligonucleotide (5'-CUCUUCAACGAG-3') was radioactively 5' end-labelled with T4 polynucleotide kinase and [γ-^{32}P]ATP using the KinaseMax™ kit from Ambion. The labelling reaction was performed with 25 pmol of RNA and 33 pmol of radioactive ATP, 1 μL of T4 polynucleotide kinase, 2 μL of the provided 10 X kinase buffer, 13 μL RNAse free water and incubated at 37 °C for one hour. After the incubation period, 1 mM EDTA was added and the reaction was halted via heating at 95 °C for two minutes. Labelled RNA was added to a reaction mixture of ScNep1 in varying concentrations from 1.2 x 10^{-4} M to 2.4 x 10^{-7} M in buffer containing 3 % glycerol, 75 mM NaCl, 2 mM $MgCl_2$, 50 mM Tris pH 7.5, and 0.1 % BME and allowed to incubate at 25 °C for one hour. After incubation, the reaction was stopped via heat denaturation and ran on a 6 % TBE polyacrylamide gel. All gels were run in triplicates, gel dried, exposed to a phosphor screen and scanned with a Storm™ Imager. Quantification of bands was performed with

Imagequant™. Figure 4.8A presents an example gel with ScNep1 concentrations showing weak binding, which conflicts with reports from *M. jannaschii* Nep1 fluorescence experiments performed by Wurm *et al*. (*22*). Upon performing the same experiment with a control RNA oligonucleotide with sequence 5'-UGUGUGUGUGUG-3' and obtaining nearly identical results, it appeared Nep1 had the ability to non-specifically bind short RNA oligonucleotides.

Fluorescence anisotropy measurements were made on the Fluorolog-3 Jobin Yvon Horiba fluorescent spectrophotometer set at 4 nm bandwidth and 1 second integration times. 5'-Fluorescein (Fl) labeled RNA (5'-Fl-CUCUUCAACGAG-3') and 3'-Fl labeled control RNA (5'--UGUGUGUGUGUG-Fl-3') was used for the binding assay. 10 nM labeled RNA was added to 1 mL of solution containing ScNep1, 75 mM NaCl, 50 mM Tris pH 7.5, 2 mM $MgCl_2$, and 0.2 % BME. A titration curve was obtained by removing 200 μL of solution and replacing 200 μL with dilution buffer containing 75 mM NaCl, 50 mM Tris pH 7.5, 2 mM $MgCl_2$, and 0.2 % BME, and 10 nM labeled RNA. The dilution buffer contained 10 nM labeled RNA to keep oligonucleotide concentration consistent throughout the entire titration curve. The reaction was done at temperature controlled 25 °C with constant stirring. Measurements were taken 8 minutes after every dilution to allow temperature and concentration equilibration with excitation at 495 nm and emission spectra averaged from 518-522 nm. The results of anisotropy reflected those seen in gel shift. Nep1 bound both the control RNA and target RNA with nearly identical affinity. Figure 4.8B shows an example of raw anisotropy data from control RNA. Anisotropy data could not be fit to an acceptable model. Both experiments described here suggest

Nep1 non-specifically binds to short RNA oligonuclotides. While, this is not ideal for measuring specific affinities of the target sequence, it does provide support that Nep1's additional positive surface can bind additional RNA (Fig. 4.2), which is expected given Nep1's environment around the 18S rRNA.

4.7: Chapter Summary

The detailed structural description of the active site and all interactions associated with the binding of the target RNA presented in the two bound structures results in an understanding of Nep1's ability to methylate pseudouridine. Through structural analysis, it is possible to present mechanistic insights including the role of a mutation causing Bowen-Conradi disease.

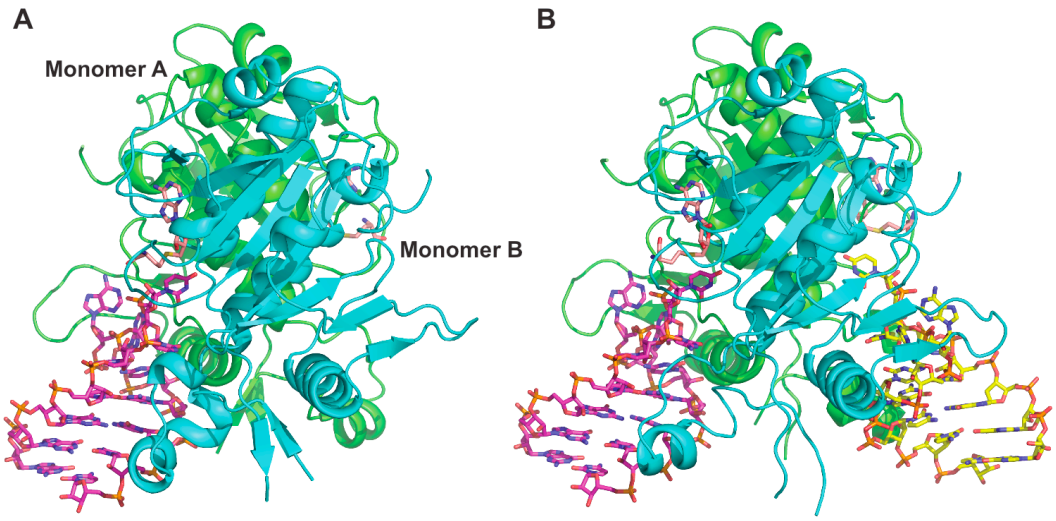


Figure 4.1: Overall Structure of the Nep1/RNA Complexes

Structures of the **A**) one RNA and **B**) two RNA complexes with SAH (pink) and RNA (magenta and yellow) shown in sticks.

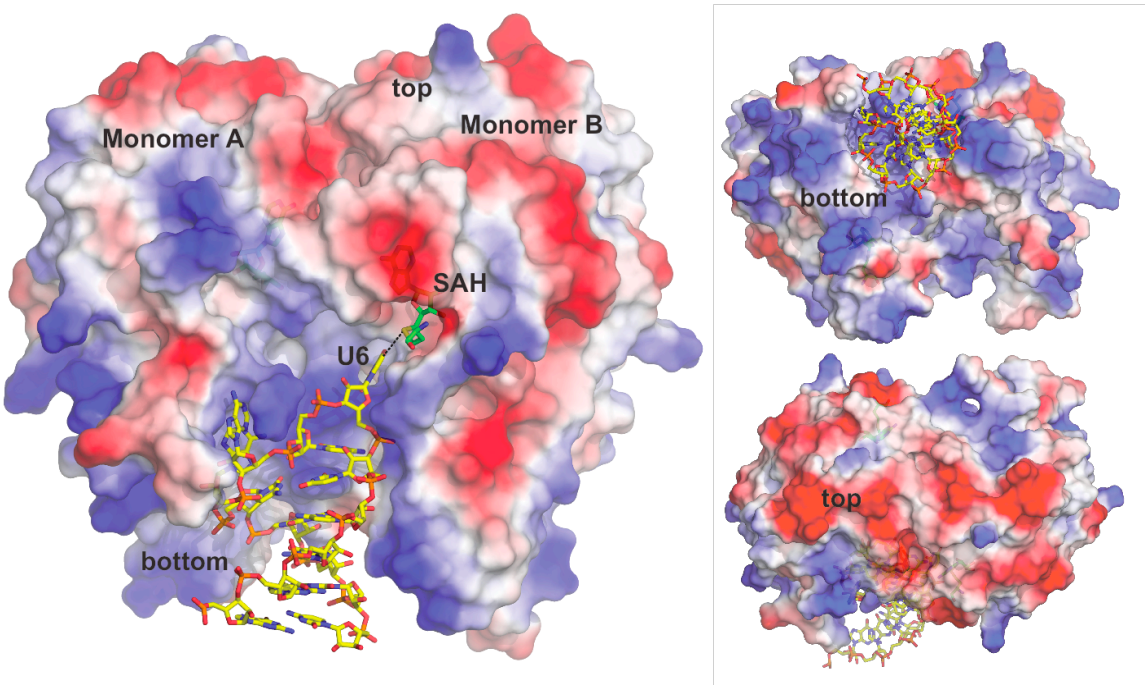


Figure 4.2: Electrostatic Surface of Nep1/1RNA Complex

Electrostatic surface of the Nep1 complex showing front, top and bottom views. The surface is colored red to blue, negative to positive charge (calculated using vacuum electrostatics in PyMOL).

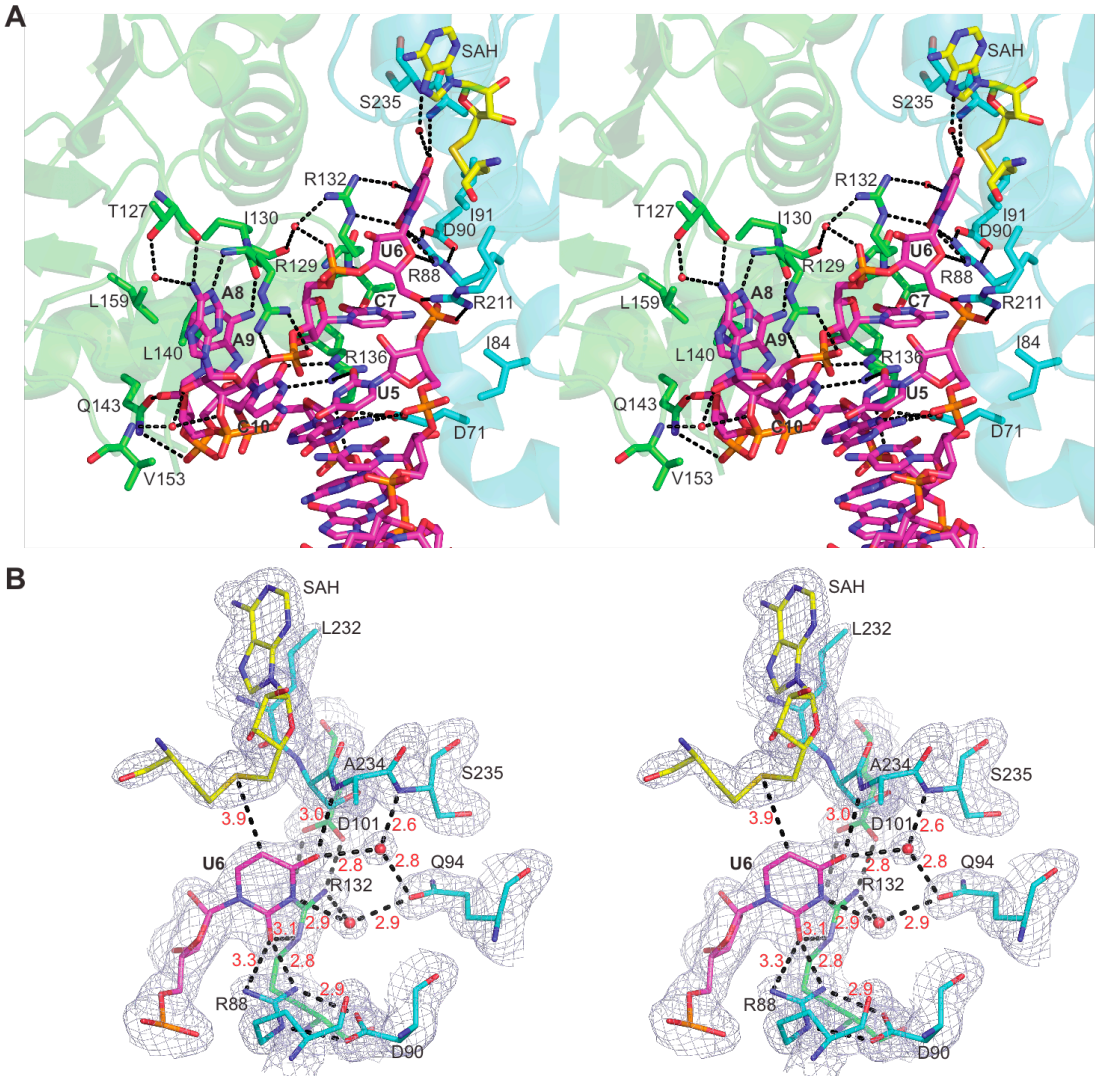


Figure 4.3: Nep1-RNA Interactions

A) Stereo figures of RNA bound in the cleft between the two monomers. Residues are colored according to monomer from which they originate (green-monomer A; cyan-monomer B). Dash lines represent connections between atoms within hydrogen bonding distance. RNA (magenta) truncated to show only the bases that contact the protein for simplicity. **B)** Stereo view of a close-up of the uridine base interactions with 2F*o*–F*c* map contoured at 1σ. Protein residues are colored as in (**A**). U_6 is shown in magenta. Distances between atoms are indicated by red text.

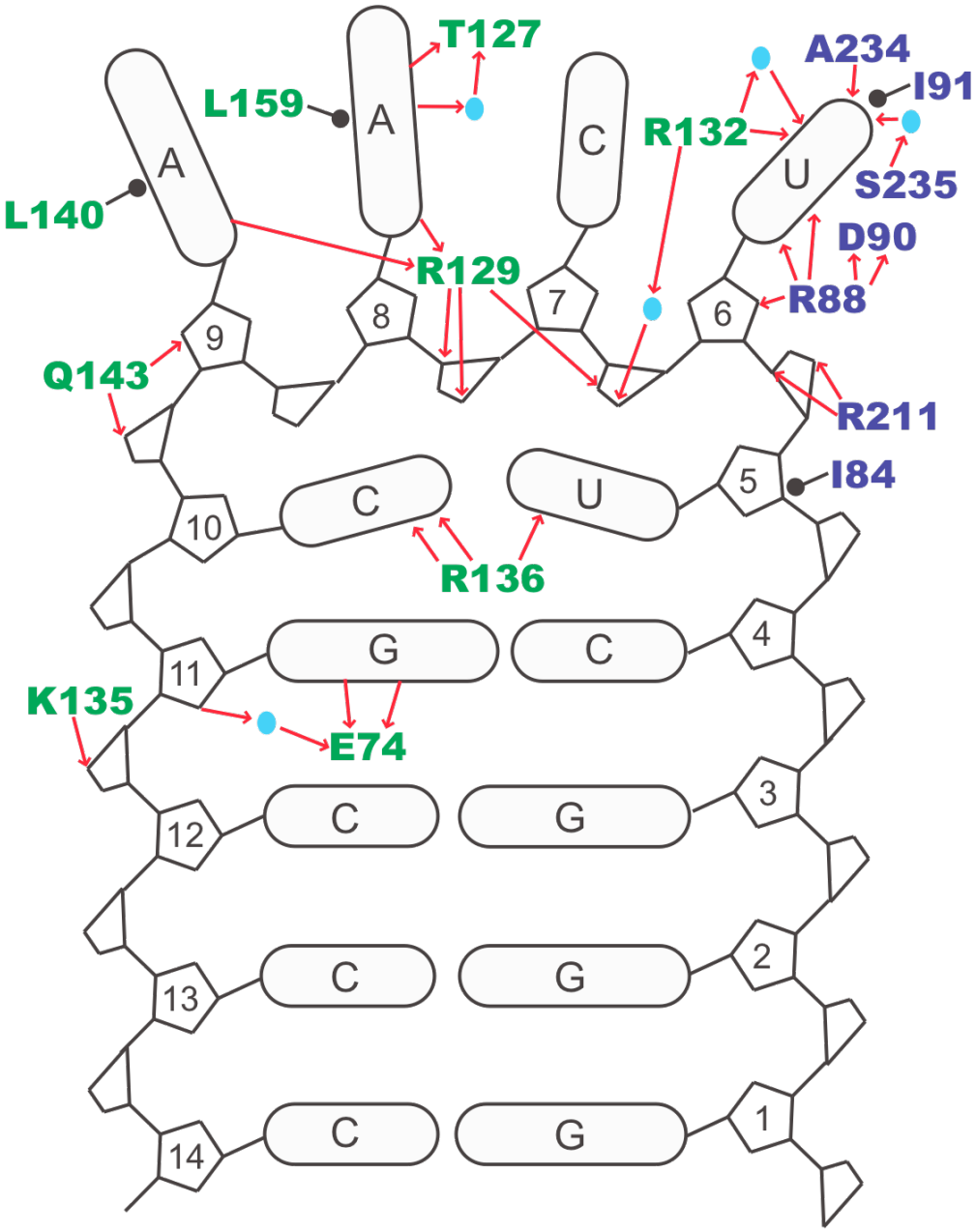


Figure 4.4: Schematic of Interactions Between Nep1 and RNA

Shown here is a diagram of the interactions between the crystallized RNA stem- loop and Nep1 residues (colored text; blue = monomer B; green= monomer A). Red arrows indicate possible hydrogen bonds and black lines terminated by black dots indicate hydrophobic contact. Blue dots indicate water molecules.

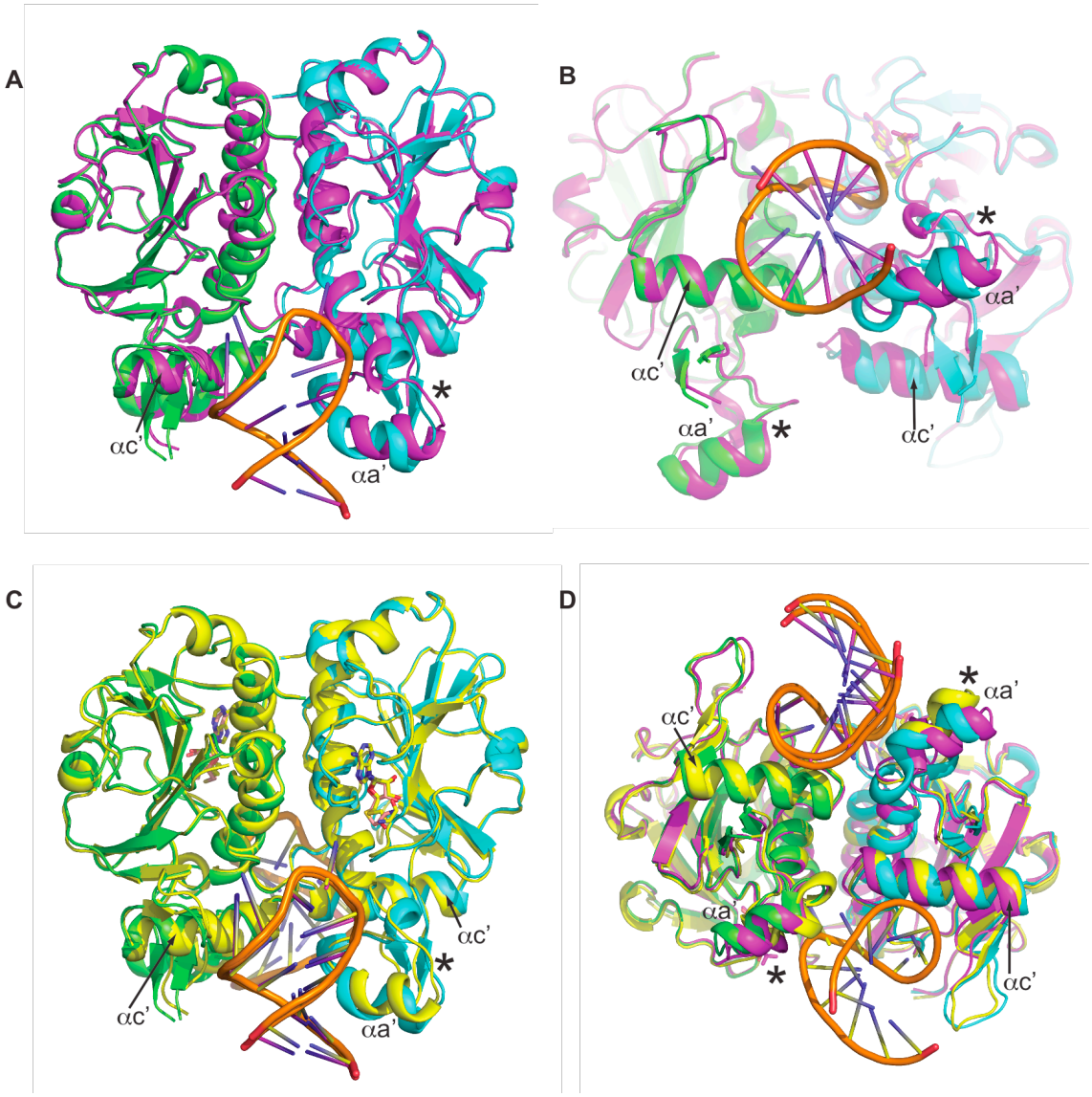


Figure 4.5: Conformation Changes Upon RNA Binding

A) Front and **B)** Bottom views of the Nep1/1 RNA complex (green-monomer A, cyan-monomer B) superposed on the free dimer (magenta). **C)** Front view of the alignment of the 1 RNA and 2 RNA (yellow) complexes. **D)** Bottom view of both RNA complexes and the free dimer. Monomer B, shown here on the right, was used to align the monomers. Asterisk indicates the Nep1 specific domain where significant structural change occurs, and αc' indicates the RNA-binding helix that shifts relative to the aligned monomer.

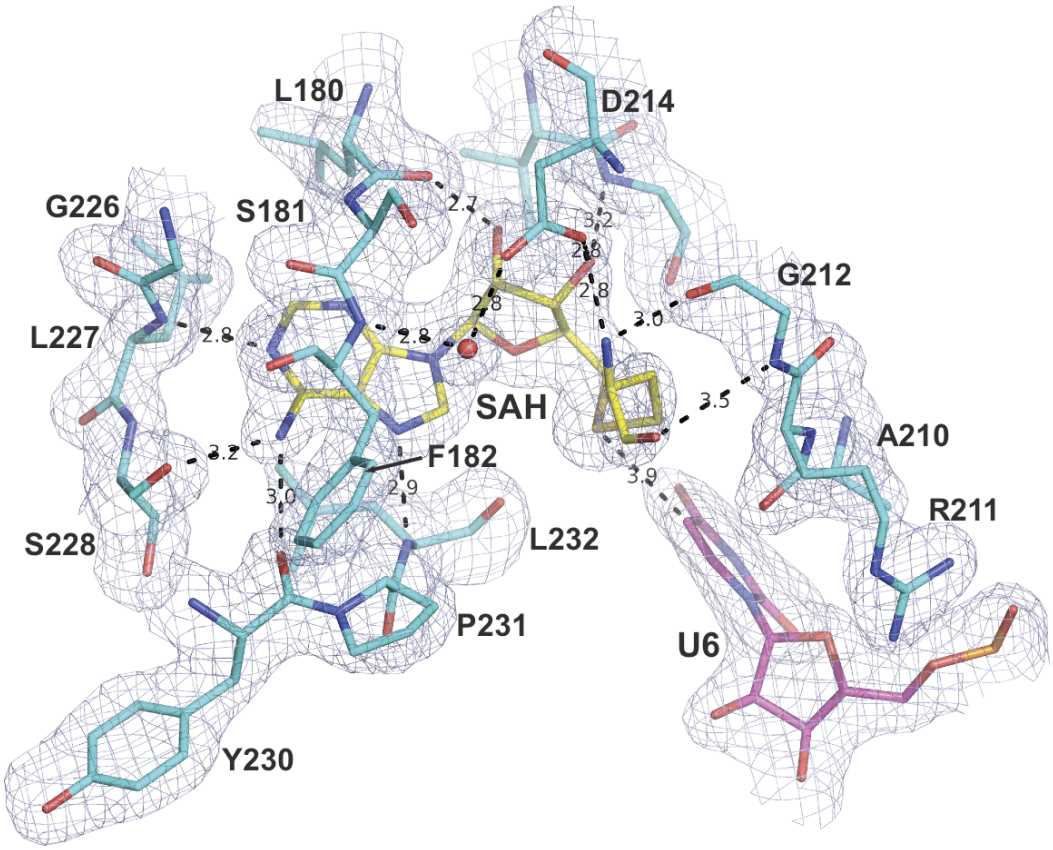


Figure 4.6: Interactions with SAH

Shown here are the interactions holding SAH (yellow) in the active site with 2Fo–Fc map contoured at 1σ. Protein residues are colored as in cyan and U_6 is shown in magenta. Distances between atoms are indicated by black text.

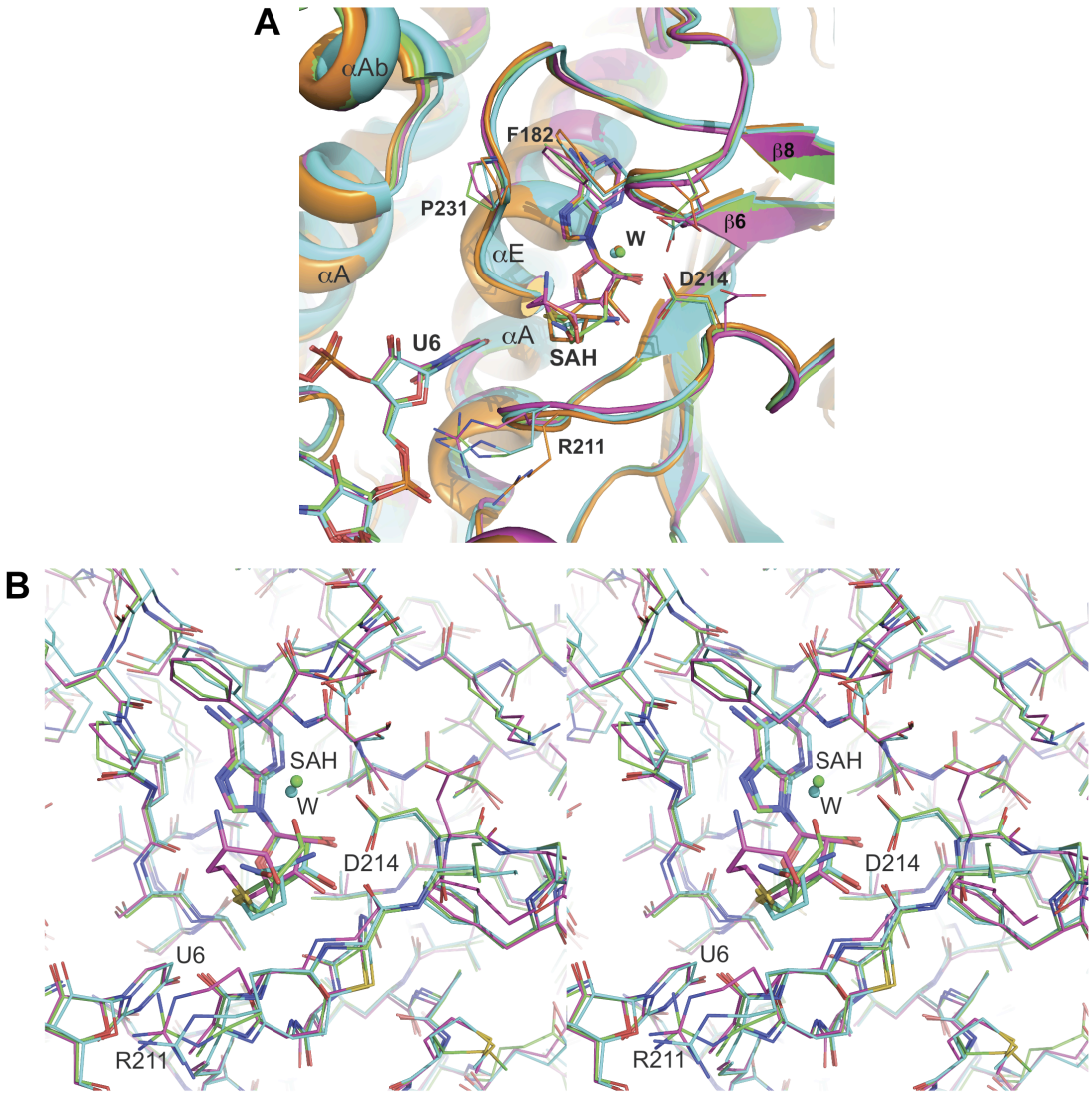
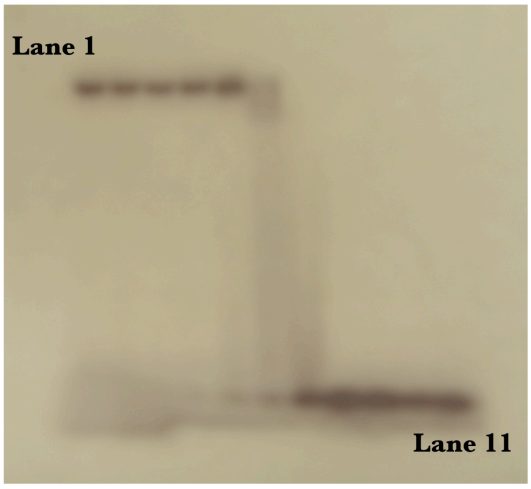


Figure 4.7: Alignment of SAM Binding Pockets

A) Alignment of the SAM-binding pockets of Nep1/1RNA complex monomer B (cyan), Nep1/2RNA monomer B (green), Nep1/2RNA monomer A (magenta) and free Nep1 dimer (orange) showing side chains of residues that show the most conformational differences between the complexes. Spheres indicate bound waters, and are labeled 'W'. **B)** Stereo view of the alignment of SAM binding sites in the Nep1/1 RNA and 2 RNA complexes. Nep1/1RNA monomer B and Nep1/2RNA monomers A and B are colored as in (**A**).

A



Lane #	ScNep1 Concentration (M)
1	0.000125
2	6.247×10^{-5}
3	3.123×10^{-5}
4	1.562×10^{-5}
5	7.808×10^{-6}
6	3.904×10^{-6}
7	1.952×10^{-6}
8	9.760×10^{-7}
9	4.880×10^{-7}
10	2.440×10^{-7}
11	No protein

B

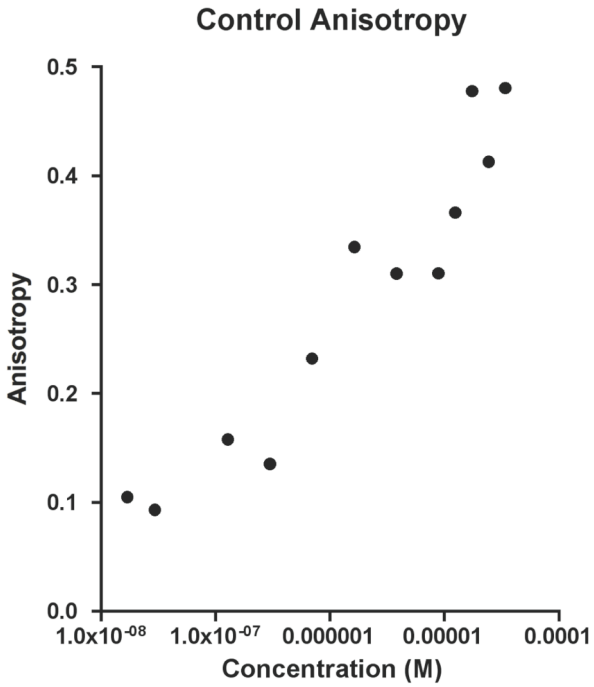


Figure 4.8: Nep1 and RNA Binding Assays

A) Storm™ scan image from phosphor screen of ScNep1 and target RNA (5'-CUCUUCAACGAG-3') electrophoretic mobility gel shift assay. B) Anisotropy data taken from Fluorolog-3 Jobin Yvon Horiba fluorescent spectrophotometer of ScNep1 binding to control RNA (5'-UGUGUGUGUGUG-Fl-3').

Chapter 5: Nep1 Discussion and Conclusions

5.1: Overview

The four deposited structures provided a leap forward in understanding the role of Nep1 in ribosome biogenesis, specifically contributing to a molecular view of the interaction with target RNA. This work supports previous research suggesting Nep1 as a pseudouridine methyltransferase along with a second function that is essential for cell survival (*21, 22*). This chapter provides discussion and conclusions based on the information gained from the analysis of AfNep1, ScNep1 free, and ScNep1 bound to one and two molecules of target RNA.

5.2: Nep1 Project Discussion and Conclusions

Prior to this work, the structures of the ScNep1 in monomeric form, and the archaeal *M. jannaschii* Nep1 dimer, allowed for the classification of Nep1 as a SPOUT methyltransferase, characterization of the Nep1 specific structural elements and analysis of the SAM binding pocket, and, in combination with mutagenesis and NMR mapping on the *M. jannaschii* Nep1, allowed for a prediction of the RNA binding site (*16, 17, 22, 78*). This report provides a view of dimeric ScNep1 in apo form and in complex with RNA, allowing a first look at a SPOUT methyltransferase bound to RNA, a high resolution analysis of the interactions of the only known N1-specific pseudouridine methyltransferase with a substrate analog, and a picture of the RNA structure that Nep1 stabilizes upon binding. These data confirm that the RNA binds to the cleft between the two monomers, which agrees with previously reported NMR mapping data and mutational analysis that reported mutation of the *M.*

jannaschii equivalents of R88, R129, R132 and R136 significantly reduced RNA binding activity (*16, 22*). In Nep1, the binding site accommodates a stem-loop structure, with the majority of binding observed between the RNA loop and the protein. Both the RNA three-hybrid screening of ScNep1 and the specificity determination on *M. jannaschii* Nep1 indicate that Nep1 binds more tightly to a predicted stem-loop structure, suggesting that this is the preferred substrate for Nep1 (*21*). However, for *M. jannaschii* Nep1, sequences that do not contain a stem can also bind (*21, 22*). We have no information about the structure of the free RNA, and no similar loops containing this sequence were found in the structure databases, but the stem-loop is predicted to form at 37 °C by RNAfold (*79*), although we are not sure if U_6 would be flipped out in the absence of protein. The RNA seen in our structures fails to contact much of the positively charged surface of the protein, leaving open the possibility that the actual binding site is larger than the selected consensus sequence.

It has been proposed that Nep1 is the first example of an N1-specific pseudouridine methyltransferase and that it catalyzes the only such modification known in 18S rRNA, the N1-methylation on the hypermodified m1acp3Ψ at position 1191 in yeast (*22*). The structural evidence, seen in the specific positioning of the C5 of uridine in the enzyme active site, supports this. The uridine is held in place by several contacts that impart this specificity, and the interaction with Arg 88 suggests that this residue plays a role in catalysis, by promoting deprotonation of the ring (Fig. 5.1). Asp 90, which is equivalent to Asp 86 that is mutated in Bowen-Conradi Syndrome, also plays a role by holding Arg 88 in place in the structure. There are no candidates for direct hydrogen binding near the C5 of the uridine in our structure that

may allow specific selection of pseudouridine, but the absence of an activating nucleophile in the vicinity of C6 in these structures rules out the possibility that this protein functions as an m5U methyltransferase, as would be suggested if uridine was the substrate (*80, 81*). In addition, m5U modification has not been observed in yeast 18S rRNA. However, this means that uridine is indistinguishable from pseudouridine in the Nep1 active site, but it will only catalyze methyl transfer on pseudouridine. This is supported by fluorescence binding data for *M. jannaschii* Nep1 reported by Wurm *et al*. that shows no significant differences in RNA binding K_d when pseudouridine is replaced with uridine in high affinity RNA binding sites (*22*).

Although it is likely that the N1-methyltransferase activity is important for Nep1 function, as supported by the severe defect caused by mutations at Asp 90, it has been reported that mutations in the SAM binding site of Nep1 that reduces its ligand binding activity by 100 fold results in no disruption of Nep1 activity (*17*). Therefore it is plausible that binding of Nep1 to the rRNA in the absence of methyl transfer activity is sufficient to carry out at least some of its role in ribosome biogenesis. In addition, overexpression of yeast ribosomal protein Rps19 compensates for deficiency in Nep1 function (*21*). An explanation put forth for these observations was that Nep1 might play a role that is independent of Nep1 methyltransferase activity (*17, 21*). A possible role may be to act as a chaperone of RNA folding during a specific assembly step. The observation of a two RNA complex in which both binding sites are occupied suggests that it is plausible that rRNA occupies both sites when Nep 1 binds *in vivo*. Based on structural analysis, the two binding sites are non-equivalent, highlighted by differences in binding to SAH

(Fig. 4.7A, B). This idea that both sites would be occupied in the ribosome is further supported by the presence of extensive positively charged surface not contacted by the stem-loops in the structures. Two of the three possible Nep1 binding sites in yeast 18S rRNA are located within 50 Å of each other in the 3-dimensional structure of the eukaryotic ribosome (Fig. 5.2A) (*30*) (*82*). Those sites are the loop terminating helix 31 and a portion of helix 42 (Fig. 1.2 p12). Based on the EM reconstruction of the eukaryotic ribosome (*30*) and the recent crystal structure of the complete yeast ribosome (*82*), in order for Nep1 to bind the site at helix 42, significant secondary structure alteration would have to occur (Fig. 5.2B; Fig 1.2 p12; Fig. 5.3). Half of the expected stem loop that is expected to form when Nep1 binds is part of helix 42. Therefore, it is possible that when Nep1 binds, there is a rearrangement of the interactions to produce the expected stem-loop (Fig. 5.3C). This rearrangement would also result in the two Nep1 sites being moved closer together, since the distance between active sites in the Nep1 dimer is ~20 Å, while it is shown to be 50 Å apart in the mature rRNA (*30*). The loop at the end of helix 42 interacts with the loop of helix 41 (Fig. 5.2B). Helix 41 is the site of most of the contacts between the rRNA and S19 (Fig. 5.2B & Fig 1.2 p12). Therefore, a possible model is that Nep1 interacts with the rRNA at two sites – the loop of helix 31, where it catalyzes the methylation of Ψ1191 and the stem of helix 42, where it acts as an RNA chaperone and breaks apart helix 42, rearranging the area and the interactions with helix 41 to promote S19 binding. In this scenario, more of the Nep1 surface would be expected to contact rRNA, as predicted by this structure. This is assuming that rRNA in early maturation stages have structures near the predicted structure of the mature rRNA. However, there is

evidence that pre-rRNA structure can differ from mature rRNA structure (*83*). In the absence of snR57, which is predicted to bind between bases 1568 and 1584, a Nep1 deletion is no longer lethal, indicating that Nep1 is only required when snR57 is present (*21*). Since snR57 guides the methylation of G1572 (*29*), a site near helix 42 (Fig. 1.2 p12), this interaction between Nep1 and snR57 may be due to structural changes that occur as a result of methylation. Another possibility, however, is that the interaction between Nep1 and its binding site in that region may promote the dissociation of snR57 from rRNA. This dissociation may be necessary for proper rRNA structure, or even for N7 methylation of G1575 catalyzed by Bud23p (*84*). In any case, the genetic interaction between the snR57 and Nep1 supports the possibility of Nep1 binding in the region near helix 42.

M. jannaschii Nep1 binds to yeast 18S rRNA sequences corresponding to the loop of helix 31 (*78*), and AfNep1 similarly is capable of binding yeast rRNA sequences corresponding to the Nep1 binding site near helix 42 (data not shown). Residues involved in key interactions with the RNA, for example Arg 88, Arg 136, Arg 129, Leu 140 and Leu 159, are identical in AfNep1. Superposition of the archaeal dimer on the yeast structures showed that the RNA interaction mode is likely to be very highly conserved between archaea and yeast (Fig. 5.4). Although these structures do not provide all the answers on Nep1 function, it does provide the basis for continued studies on the ancient structural mechanisms utilized by Nep1 in rRNA processing.

5.3: Hypothesis of Nep1's Role in Ribosome Biogenesis

Taking all the information presented, a hypothesis for Nep1's role in ribosome biogenesis can be suggested. For Nep1 to bind both target sites in helix 31 and helix 42, there must be a major conformation change in helix 42 (Fig. 5.5) bringing the two active sites closer together. I propose Nep1's main role in ribosome biogenesis is to facilitate this conformational change and chaperone the folding of the immature pre-rRNA into the essential structure needed for the loading of ribosomal protein S19 to helix 41, which is in direct contact with helix 42 in the mature 18S rRNA. Furthermore, I propose Nep1 preferentially binds helix 31 prior to the target sequence of helix 42. This will set Nep1 in position to dissociate snr57, making room for Nep1's interaction with helix 42. Finally, after S19 is properly loaded, Nep1 is disassociated from the ribosome caused either by the rearrangement of helix 42 as seen in the final 18S rRNA structure.

5.4: Future Directions

In order to further understand the role of Nep1 in ribosome biogenesis, questions persist surrounding the proposed chaperone function. The ScNep1/1RNA and ScNep1/2RNA complex structures shows that Nep1 can bind RNA simultaneously on both sides of the dimer, but its still unknown if this does indeed happen simultaneously. Does Nep1 preferentially bind one target before the other or do both sites bind at the same time? Next, Does Nep1 direct a major conformational shift in helix 42 that results in the dissociation of snr57 and the subsequent loading of S19? To look at these questions, I believe work must shift from short RNA oligonucleotides to longer sections of the 18S rRNA that encompasses both Nep1

target sites and the location of S19 binding. Utilizing a longer section of the 18S rRNA in different binding assays that includes structural biology from both x-ray crystallography and NMR, RNAse secondary structure foot printing and fluorescence binding assays, could provide needed answers to these important questions. Finally, small angle x-ray scattering (SAXS) is currently an emerging field and may provide an ideal technique in the near future for visualizing the proposed rearrangements of the 18S rRNA.

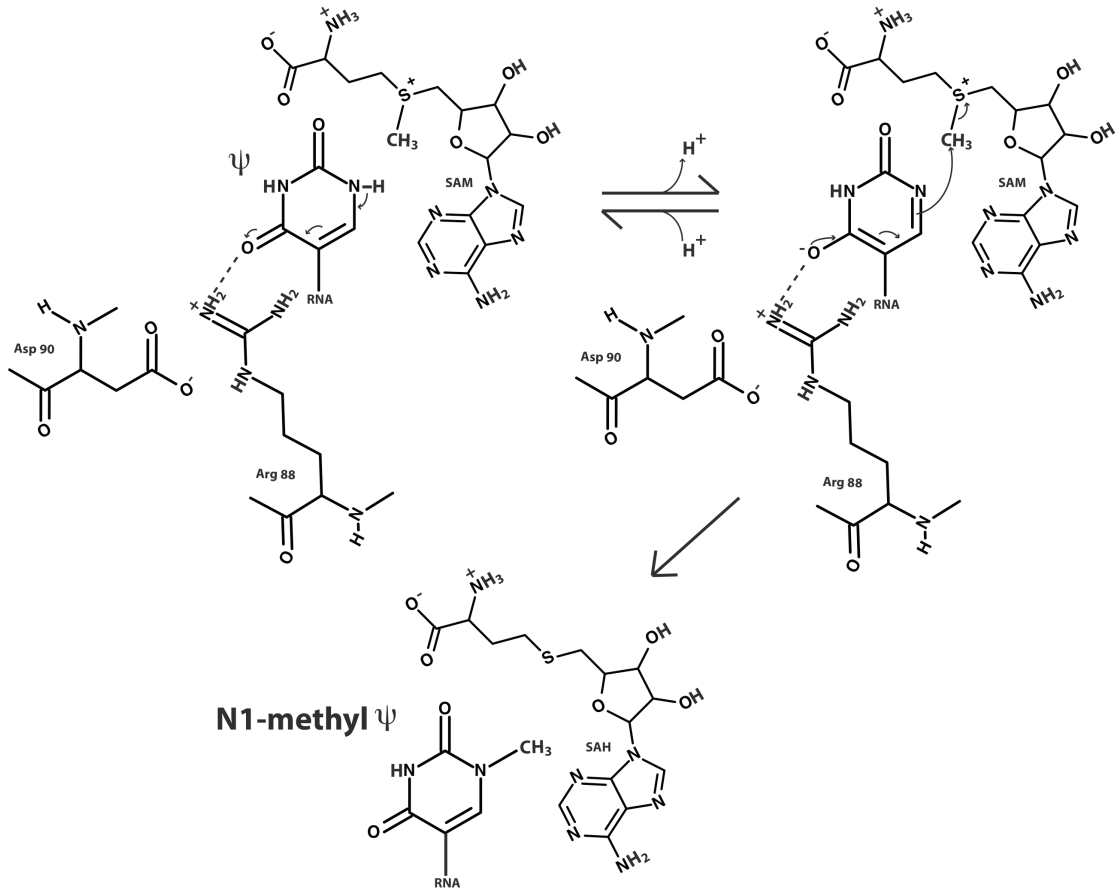


Figure 5.1: Proposed Nep1 N1-pseudouridine Methyltransferase Mechanism

Putative mechanism for the transfer of a methyl group to the N1-position of pseudouridine (ψ) by Nep1.

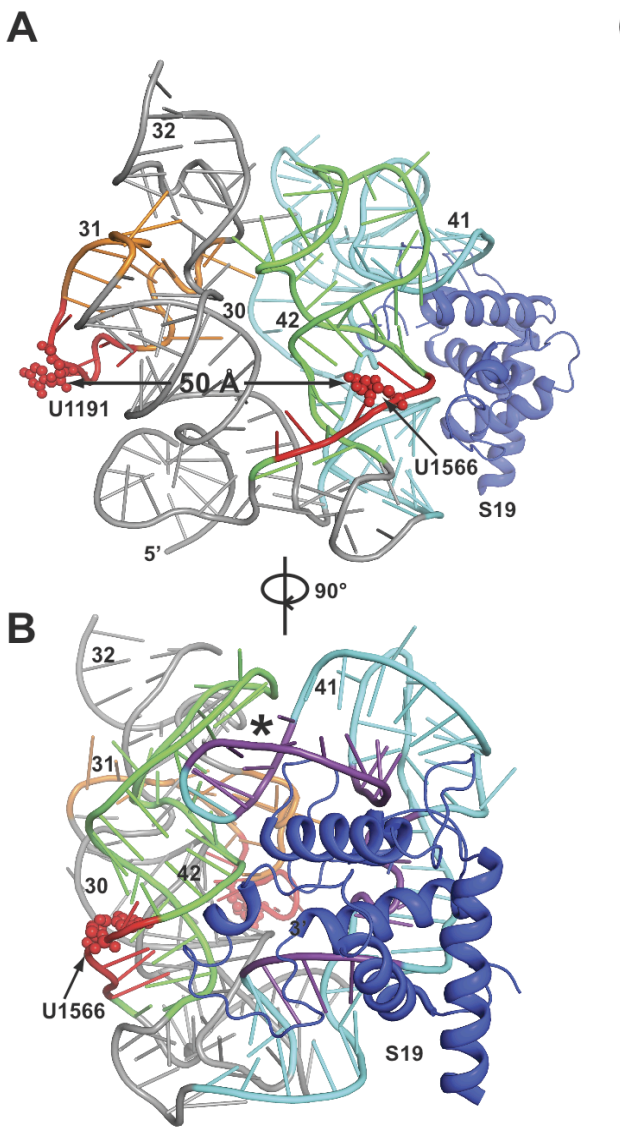


Figure 5.2: Possible Role for Nep1 in S19 Loading

A) and **B**) Three dimensional structure of the head region of eukaryotic 18S rRNA, truncated to remove bases 1212–1439 for simplicity. (**A**) and (**B**) were generated from PDB ID 3JYV [cryo-EM model of the 40S subunit of the *Thermomyces lanuginosus* ribosome (*30*)].

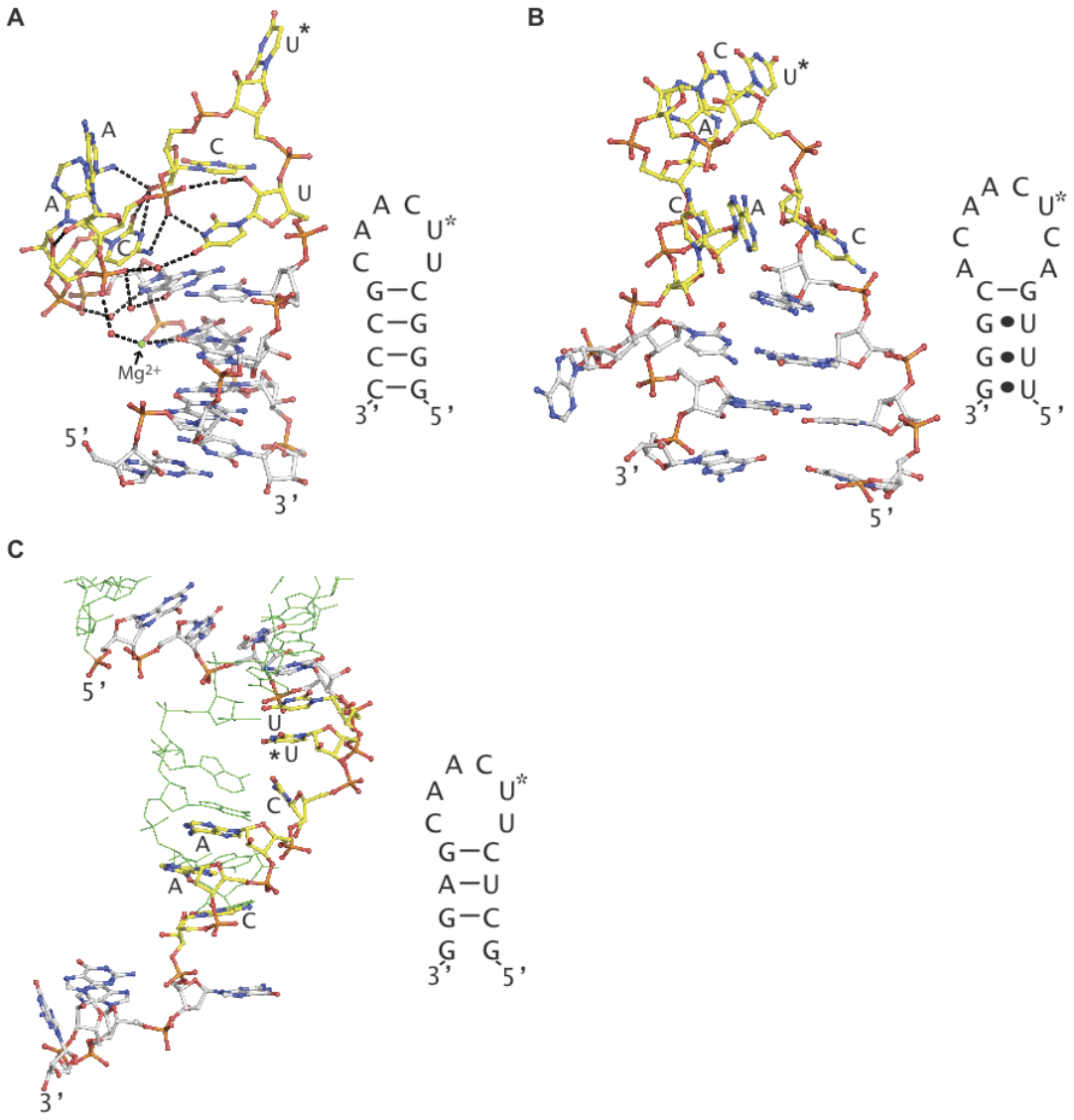


Figure 5.3: RNA Structures in Nep1 Binding Sites

A) Structure of RNA in Nep1 complex with schematic of the secondary structure shown on the right. **B)** Structure of Nep1 site at helix 31 and **C)** helix 42 (from the cryo-EM reconstruction of the eukaryotic ribosome (*30*)). Secondary structure predicted to be induced upon Nep1 binding shown on the right. For all figures, the consensus binding site is colored yellow and the * indicates the uracil inserted in the Nep1 active site.

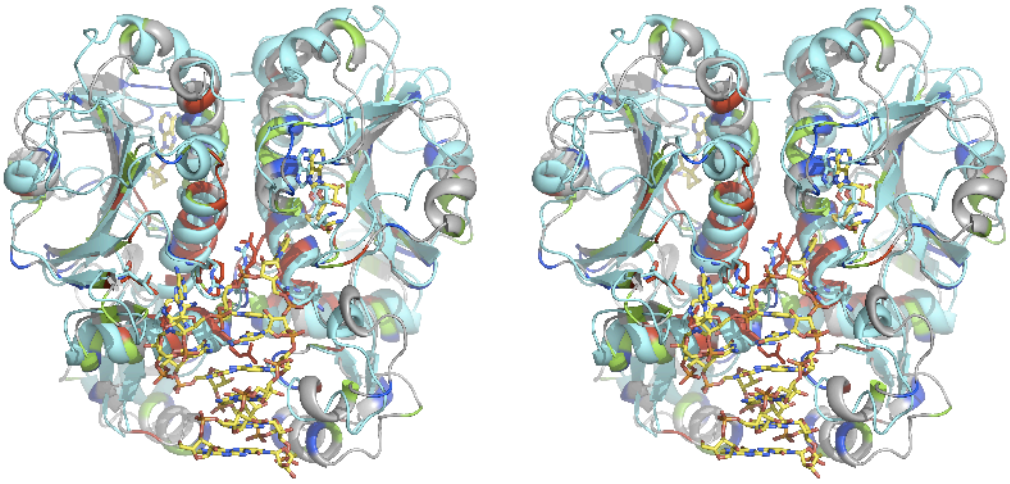


Figure 5.4: AfNep1 and ScNep1 Aligned with Sequence Conservation

Stereo view of ScNep1 (cyan) bound to one RNA molecule aligned to AfNep1 (with symmetry mate to form dimer), colored to show conservation of the RNA binding site between archaeal and eukaryotic Nep1 proteins (red = identical; green = highly conserved; blue = weakly conserved). Sequence alignment done using ClustalW (*46*).

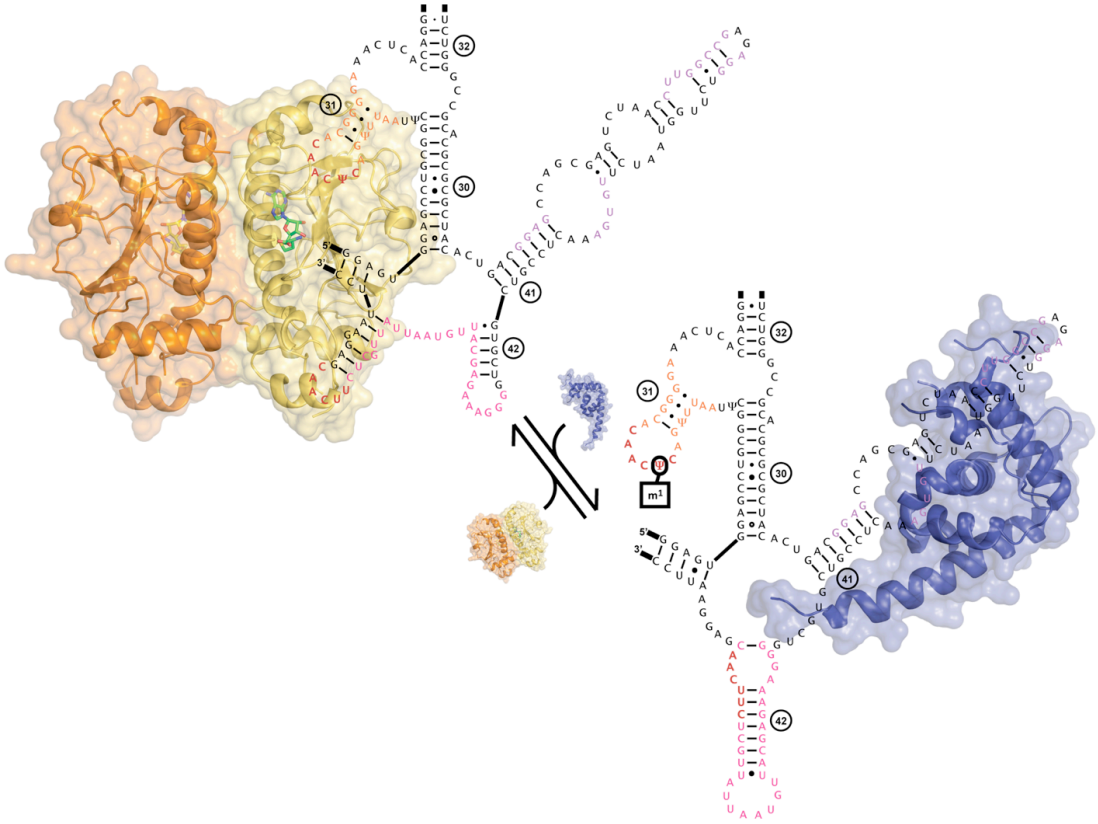


Figure 5.5: Proposed Chaperone Role of Nep1 in Ribosome Biogenesis

Model of the proposed role of Nep1 in ribosome biogenesis highlighted by Nep1's ability to bind both target sequences simultaneously and chaperone the folding of helix 42, which will allow the loading of S19 to helix 41.

Chapter 6: Purification of Rio3

6.1: Overview

Rio3 is an atypical kinase with an unknown function and unknown structure. By sequence analysis with other family members, Rio3, known as RioK3 in humans, possesses a conserved RIO domain and an extended N-terminal domain of approximately 200 amino acids (*35, 37, 39, 48*). The goal of this project was to solve the crystal structure of the RIO domain for RioK3. The full-length construct (519 residues) suffered from insolubility problems when expressed in *E. coli*. Therefore, the focus of research was spent trying to express, purify, and crystallize a shortened construct containing only the RIO domain of RioK3 (residues 193-519). Presented in this chapter is the best purification protocol for the RIO domain of RioK3 and for the N-terminus of RioK3. Initial crystallization trials from the resulting protein acquired from this purification protocol are also described in this chapter.

6.2: Purification of RioK3

The human shortened RioK3 gene (residues 193-519) with an N-terminal TEV cleavage site was PCR amplified using sequence specific primers (5'-GAGAACCTGTACTTCCAGGGAATGGATTTAAAACTATCAAAC-3' and 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTTATCATTCATCATATAGTAGT GGTGGGTC-3') and recombined into pDONR201 utilizing Invitrogen's Gateway™ technology. The resulting plasmid was sequenced and confirmed the correct sequence for the shortened RioK3 construct with N-terminal TEV cleavage site. The plasmid was recombined into pDest527 designed with an N-terminal 6X-His tag (plasmids

obtained from Protein Expression Laboratory, SAIC, Frederick, MD). RioK3-Short construct was transformed into *E. coli* Rosetta BL21-DE3-pLysS cells via heat shock. The construct was expressed in 1 L LB broth containing ampicillin (100 μg/ml) and chloramphenicol (34 μg/ml) at 37 °C until the optical density of the cells reached 0.5 (approximately 4 h). At the correct cell density, the cells were transferred to 20 °C incubator/shaker and allowed to cool before induction. The RioK3-Short construct was induced with 1 mM isopropyl thio-β-D-galactoside (IPTG) and grown overnight (16 h) at 20 °C, shaking at 250 rpm. Cells were harvested by centrifugation and cell pellet was frozen overnight at -80 °C. The frozen pellet was resuspended in 50 mL lysis buffer containing 500 mM NaCl, 50 mM Tris pH 8.0, 0.2% β-mercaptoethanol (BME), 2 mM $MgCl_2$, 1 mM adenosine triphosphate (ATP), 0.1 mg/ml DNase1 (Roche), 10 μg/ml RNaseA (Roche), and 0.1 X Bugbuster (Novagen) per liter of culture. The resuspended pellet lysed on ice after stirring for 1 h. The lysate was transferred to centrifuge tubes and the cell debris was spun down at 17000 rpm in Beckman preparative (rotor: Beckman 45 Ti) ultracentrifuge for 40 minutes at 4 °C. The supernatant was collected through a 0.22 μm filter. The supernatant containing the soluble RioK3-Short construct was loaded onto an equilibrated 5 mL GE Healthcare HisTrap HP column using the BioRad Duoflow FPLC. The column was equilibrated with buffer A containing 500 mM NaCl, 50 mM Tris pH 8.0, 0.2% BME, 2 mM $MgCl_2$, 1 mM ATP. The bound RioK3-Short construct was washed extensively with buffer A (>10 column volumes) followed by another extensive wash including 100 mM imidazole. RioK3-Short was eluted from the column using a linear gradient from 100 mM imidazole to 1 M imidazole over ten column volumes.

The fractions containing RioK3-Short were combined and set up in dialysis membrane with buffer A and 2 mg TEV protease. The cleaved protein was loaded back onto the 5 mL HisTrap column. The RioK3-Short protein did not come through in the flowthrough: 20 mM imidazole in buffer A was needed to elute the bound protein. The fractions containing RioK3-Short protein were pooled together and concentrated to a final volume of 3 mL. A Superdex 200 16/60 size exclusion column was equilibrated with buffer A and the Rio3-short protein was passed over the column as a final purification step. The fractions containing purified RioK3-Short protein were combined and concentrated to 15 mg/ml for crystallization. All steps of purification were analyzed for purity by SDS-PAGE (Fig. 6.1). Table 6.1 shows the RioK3 constructs that were expressed for purification and crystallization trials.

6.3: Initial Crystallography of RioK3-Short

The purified RioK3-Short construct was concentrated to 15 mg/mL and immediately screened for crystals using the Art Robbins Phoenix™ liquid handler. The sitting drop 96 well 3-well Intelliplate was chosen for screening with conditions known to promote crystallization. These varying conditions are combined into the 96 deep-well format and were purchased from Qiagen (PEG™ Suite, Cryos™ Suite), Emerald Biosystems (Wizard 1™, Wizard 2™, Wizard 3™, Natrix™) and Hampton Research (Index™). The 3-well format of the intelliplate allowed for a concentration range for each condition by setting up the sitting drops in 1:2, 1:1 and 2:1 protein:well solution ratios. In total, the initial crystallization screen for RioK3-Short consisted of 1,152 different drops. Of those 1,152, one condition yielded extremely small (~5x30μm) protein crystals shown in Figure 6.2A. The crystals grew in 25% PEG-

3350, 0.1 M HEPES pH 7.5, 0.2 M $MgCl_2$. Those crystals were reproducible from protein expressed in minimal media (described in chapter 7) via the hanging drop method (Fig 6.2B) and sent to the NECAT APS beamline at Argonne National labs for testing of diffraction. This work was promising, yielding diffraction to ~5.5 Å shown in Figure 6.3.

6.4: Expression and Purification of the N-Terminus of RioK3

While working on the purification and crystallization of the RioK3 RIO domain, the N-terminus of RioK3 was shown to bind to polyubiquitin (*85*). N-terminal fragment 16-153 was shown to bind to Lys-48 linked and Lys-63 linked Ubuitin (*85*). RioK3 does not possess a known Ubiquitin binding domain, which suggests RioK3 could bind Ubiquitin via a novel mechanism. Using Invitrogen Gateway™ cloning, RioK3-1-153 and RioK3-1-193 were cloned with an N-terminal histidine tag followed by a TEV cleavage site with forward sequence (5'-GAGAAC CTGTACTTCCAGGGCATGGATCTGGTAGGAGTGGCATCG-3' and reverse primers 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTTATCACTTTTTAGG AGTGGGAACCGGTTT-3' and 5'-GGGGACCACTTTGTACAAGAAAGCTGGG TTATCAAATTCCATCTCCTACCTGAAACTC-3', respectively. After unsuccessful purification attempts due to insolubility, RioK3-1-193 was recombined into pDest566 with maltose binding protein (MBP) as an N-terminal fusion protein resulting in HIS_6-MBP-TEV-(RioK3-1-193). The fusion protein was used to improve solubility. The expression and purification scheme was similar to RioK3-Short with two modifications. First, the lysis and size exclusion buffers contained 200 mM NaCl, 50 mM Tris pH 8.0 and 0.2 % BME. The elution buffer contained 200 mM NaCl, 50

mM Tris pH 8.0, 0.2% BME and 1 M imidazole. Second, after TEV cleavage MBP-RioK3-1-193 did not require 20 mM imidazole for elution and came out as expected in the flowthrough. All steps of purification were analyzed by SDS-PAGE and shown in Figure 6.4. The successful purification of the N-terminus of RioK3 provides a great starting point for the next researcher to look at interactions with Ubiquitin.

6.5: Chapter Summary

The initial crystallography results showing diffraction from crystals acquired from the purification protocol presented in this chapter was very promising. The crystals were reproducible and they diffracted! This beginning success provided a great starting point for crystal optimization.

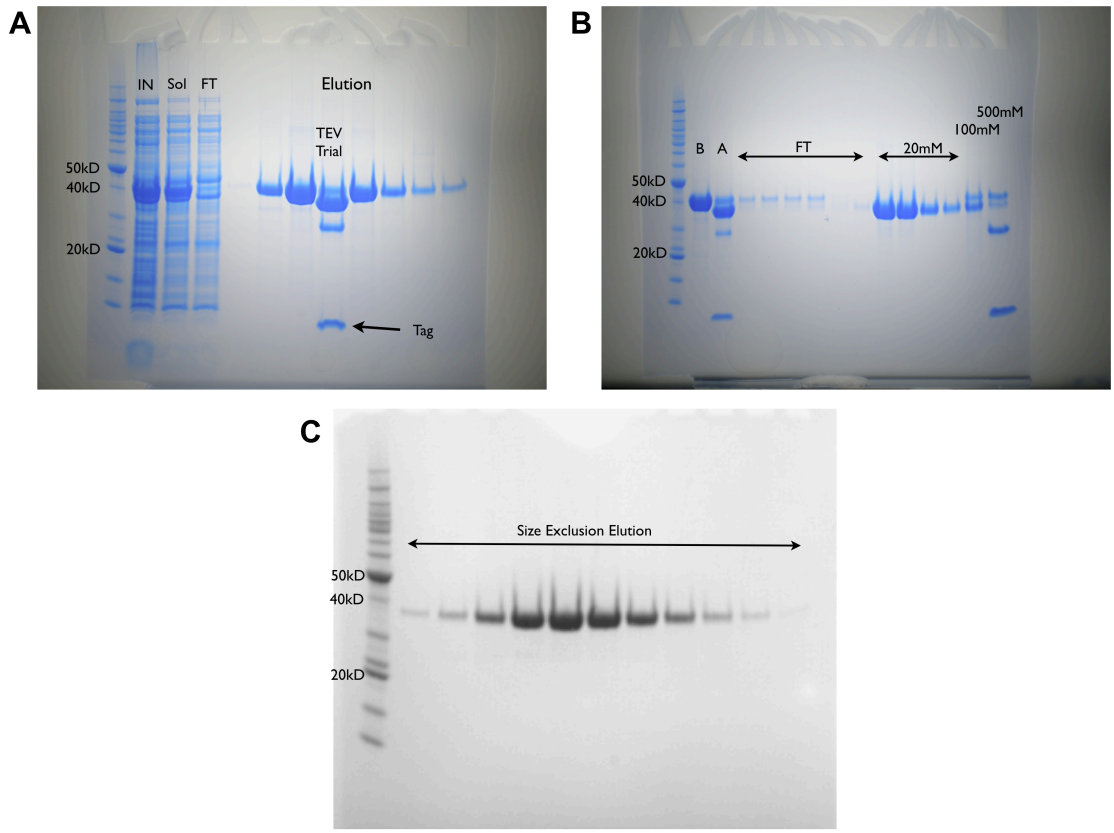


Figure 6.1: RioK3-Short Purification Gels

A) SDS-PAGE of RioK3-short after first pass over GE Healthcare HP affinity column (IN=induced sample, Sol=supernatant after lysis, FT=flowthrough, TEV Trial=20uL elutant incubated with 20 μL TEV protease (1 mg/mL) for 30 minutes on rocker and ran as lane 8, Tag=histidine tag cleaved during TEV trial). **B)** RioK3-Short after cleavage passed over affinity column to remove tag and uncleaved protein (B=before cleavage, A=after cleavage). Concentrations indicate amount of imidazole required for elution from column. **C)** Elution from GE Healthcare Superdex 200 16/60. Expected MW of Riok3-Short is 37,717 Da.

Table 6.1: RioK3 Expressed Constructs

Construct	**Soluble (Y/N)**	**Purified (Y/N)**	**Crystallized Y/N (Å)**
RioK3-Full Length	Y	Y	N
RioK3-Short (193-519)	Y	Y	Y (5.5 Å)
RioK3-S199A-Short (193-519)	Y	Y	Y (8.0 Å)
MBP-RioK3-Short (193-519)	Y	Y (80%)	N
RioK3-KK223AA (193-519)	Y	Y	Y
RioK3-EKK222AAA (193-519)	Y	Y	Y
RioK3-KE277AA (193-519)	Y	Y	N
RioK3-EKE276AAA (193-519)	Y	Y	N
RioK3-193-474	N	N/A	N/A
RioK3-248-474	N	N/A	N/A
RioK3-248-519	N	N/A	N/A
RioK3-1-142	N	N/A	N/A
RioK3-1-153	N	N/A	N/A
RioK3-1-193	N	N/A	N/A
MBP-RioK3-1-193	Y	Y	N

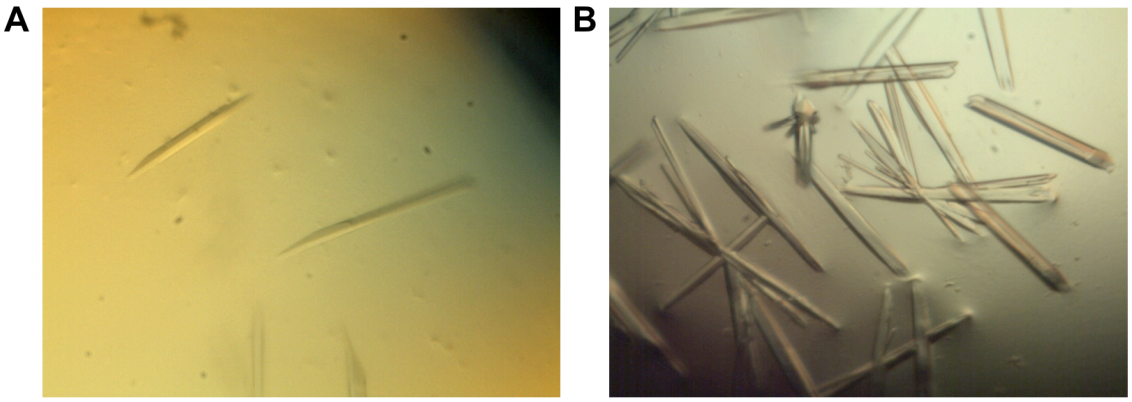


Figure 6.2: RioK3-Short Crystals

A) First Riok3-Short crystals obtained from G12 of the Index™ screen in the 2:1 protein:screen ratio**. B)** Reproduced crystals from RioK3-Short grown in minimal media. Diffraction of crystals from (**B**) screened at the synchrotron resulted in diffraction to ~5.5 Å.

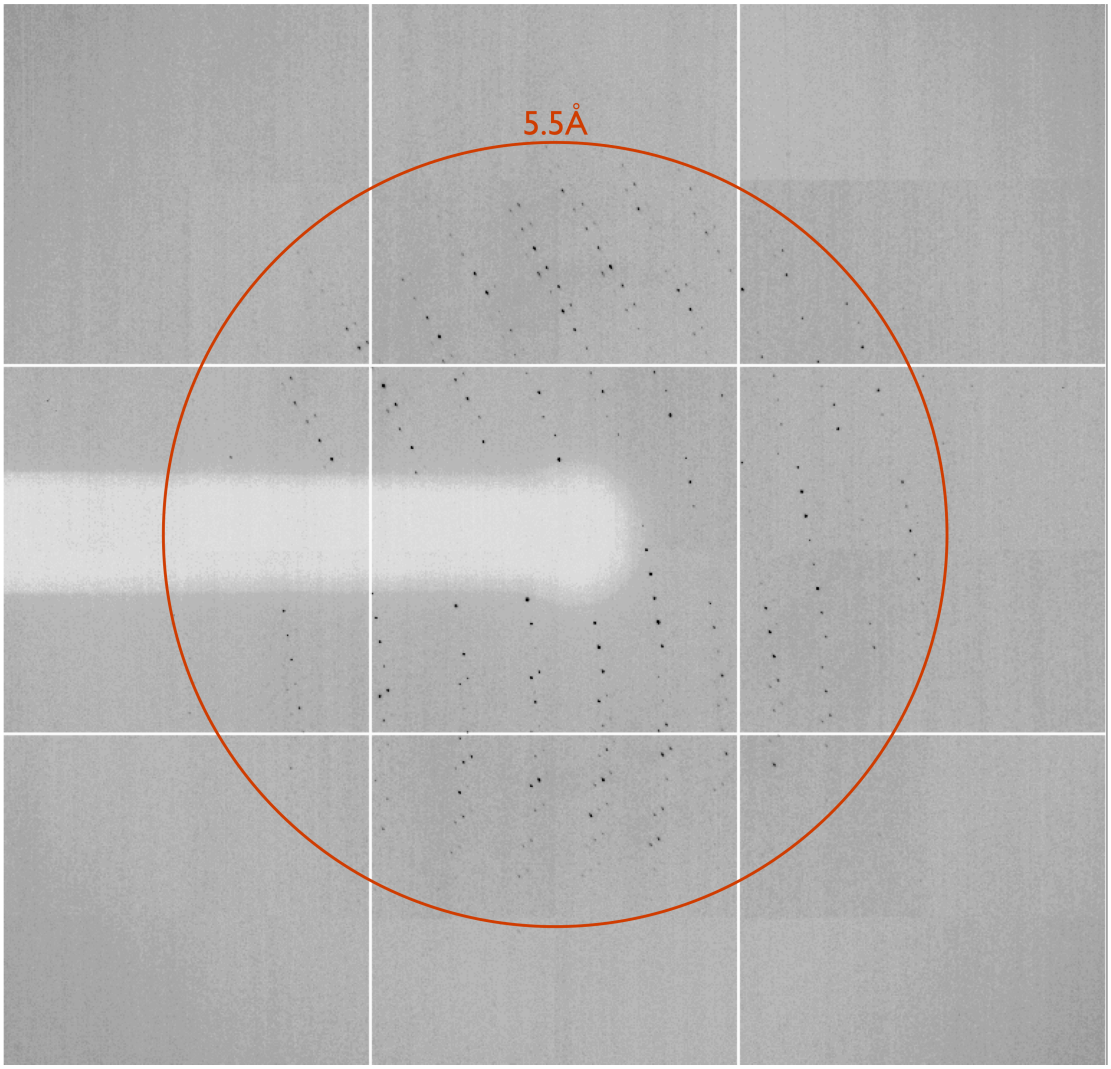


Figure 6.3: Diffraction of RioK3-Short

Diffraction image of a RioK3-Short crystal grown in minimal media, collected at NECAT APS beamine in Argonne, Illinois. Red circle represents 5.5 Å resolution.

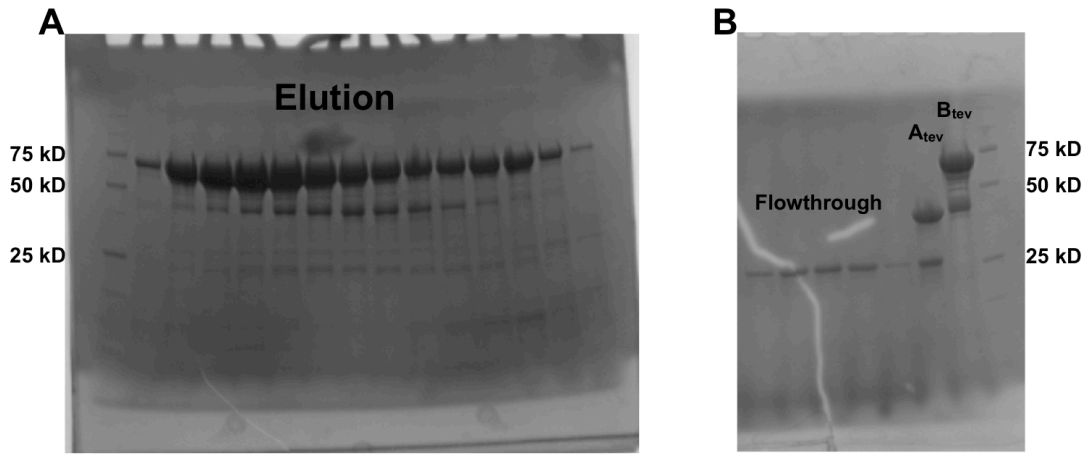


Figure 6.4: MBP-RioK3-1-193 Purification

A) SDS-PAGE of MBP-RioK3-1-193 after first pass over GE Healthcare HP affinity column B) Cleaved RioK3-1-193 was passed over the affinity column a second time and collected in the flowthrough. This gel shows effiicient cleavage after TEV. (B_{tev}=before cleavage, A_{tev}=after cleavage).

Chapter 7: Optimization of RioK3

7.1 Overview

The crystals acquired from the RioK3-Short construct and purification protocol presented in chapter 6 provided the first look at diffraction. The 5.5 Å resolution diffraction however, does not provide enough resolution to acquire a solution, either by *de novo* phasing or by molecular replacement. This chapter will describe the many methods used to optimize RioK3 in order to acquire strong enough diffraction for structure solution.

7.2: General Optimization Methods

7.2A: Precipitant, pH, Temperature, Salt, Protein Concentration/Ratio Optimization

The first method of optimization is always setting up conditions that vary the original crystal condition discovered during initial screening, in this case Index™ G12. For RioK3-Short, this general optimization method was performed methodically varying the precipitant concentration (16-30% PEG-3350), buffer pH (0.1 M HEPES pH 5.5-9.0), and salt concentration (0.1-0.4 M $MgCl_2$) independently of each other in a 24-well hanging drop Qiagen Cryotool (crystal plate). Also, concentration of RioK3-Short ranged from 8-30 mg/ml, as crystal size and quality can vary greatly depending on protein concentration. The temperature of crystal tray incubation was performed at 20°C or 4°C. The goal of the optimization was to acquire bigger, more defined crystals. However, these methods of optimization did not work. There was an

excess of precipitation in the hanging drops in nearly every condition. What was causing this precipitation and lack of crystal formation? After trial and error and multiple rounds of purification and optimization trays, the reason for this instability became clearer. RioK3-Short was unstable and precipitated when diluted from its original purification buffer containing high salt (500 mM NaCl), high ATP concentration (1 mM) and 2 mM $MgCl_2$. To counteract this dilution, I chose to develop a method of setting up drops that would create a range of dilution (Fig 7.1A). An "X" of concentrated protein (4 μL) was created using a pipette tip. At the center of the formed "X", 1 μL of precipitant was added, this is depicted in Figure 7.1A. If the dilution of original condition were the cause of the instability, then this "X" method would show heavy precipitation in the middle, crystals in the precipitation interface (also known as the metaphase), and finally soluble protein without crystals towards the edges of the "X". As seen in Figure 7.1B, this method confirmed RioK3-Short becomes unstable as the original conditions are diluted. Also, this "X" method of setting up drops, yielded many clean crystals and provided reasoning to set up unconventional ratios of protein:precipitant that was later optimized to 6:1. Although bigger, cleaner crystals were obtained by general optimization, the diffraction was limited 6.0 Å.

7.2B: Minimal Media/ Minimal Media with Se-Met

Before changing the construct and introducing mutagenesis, another form of optimization is to modify the method of protein expression. This can be done through expression in minimal media and minimal media containing selenium methionine (Se-Met). Along with experience seen in the lab, trends in literature and recognizing

examples at conferences, there is ample supporting evidence that expression in minimal media can improve crystal quality and diffraction (*86, 87*). It is still unclear and not yet in literature what is the cause of this improvement, but the current hypothesis is the minimal media slows expression and promotes more precise protein folding resulting in a more homogenous protein for better, tighter crystal growth. The presence of a strong diffracting heavy atom (Se-Met), introduced during expression, may also improve diffraction. RioK3-Short was grown in minimal media and also in minimal media containing Se-Met. Bigger cleaner crystals were obtained in minimal media (Fig 6.2B) and resulted in the 5.5 Å diffraction pattern shown in chapter 6 Figure 6.3. The diffraction pattern was indexed with DENZO, part of the HKL2000 program suite, where the point group and unit cell dimensions were determined (Fig 7.2A, B) (*56*). The unit cell was enormous with dimensions of a=142.0, b=142.9, c=273.5 in the primitive hexagonal point group later determined to be space group P32 during scaling with Scalepack, which is also packaged into HKL2000 (*56*). Table 7.1 shows the data logfile from Scalepack from a near complete dataset collected from a RioK3-short crystal (Fig 6.2B) grown in minimal media.

Although high symmetry is present in the crystal, the Matthews Coefficient (MC) calculation suggests between 13-23 (most likely 18) molecules in the asymmetric unit (*88, 89*). The MC program uses the space group and unit cell dimensions to calculate a total volume in the cell. After calculating the total volume, the program then uses the expected molecular weight of the protein and predicts the percentage of solvent in the asymmetric unit, or in other words, the specific volume (Matthews Coefficient) contributed from the inputted protein molecular weight. Table

7.2 shows Matthews Coefficient output logfile with the number of monomers possible in the asymmetric unit ambiguously between 13-23 monomers corresponding to a solvent content from ~35%-65%. The most common percentage of solvent found in the PDB database is 47% solvent, which corresponds to 18 or 19 monomers for RioK3-Short (*89*). The oversized unit cell and low resolution diffraction combined with calculations to predict the large number of monomers in the asymmetric unit suggests RioK3-Short crystals were disordered and not packed very tightly. The goal of optimization from this point was to shrink the unit cell and promote tighter crystal packing.

7.2C: Crystal Manipulation (Annealing, Dehydration, Seeding)

Crystallography techniques that involve physical manipulation of already grown crystals include annealing, dehydration and seeding. Annealing of crystals is used to eliminate possible disruptions in crystals lattices caused by freezing (*90, 91*). Cryogenic conditions are supplemented to crystal conditions to protect crystals during freezing, which is required before shooting with high energy x-rays; however, sometimes even in the presence of a suitable cryogenic condition freezing still causes deformation and disorder in the crystal. Specifically, freezing can cause an expansion of the solvent in the crystal and cause disordered crystal packing resulting in poor diffraction. Annealing is the process of warming the crystal to room temperature after flash freezing. Two types of annealing were performed with Riok3-Short crystals. The first, known as annealing on the loop, consists of blocking the cryostream while the crystal is mounted on the goniometer just long enough to thaw and then flash refreeze by unblocking the cryostream (*91*). This method destroyed

nearly all diffraction; leaving just smeared diffraction, which are common in severely damaged crystals. The second, known as macromolecular crystal annealing, was performed by removing a crystal from the cryostream and placed back into the crystal condition containing 20% ethylene glycol for five minutes (*91*). After sitting for five minutes, the crystal is then looped again and flash frozen. This method of annealing also did not improve diffraction.

Dehydration is performed for a completely different reason than annealing. Dehydration is used to remove water molecules from the crystal in order to shrink the unit cell (*91, 92*). This method was very appealing for RioK3-Short because of the enormous unit cell observed in previous diffraction data. Dehydration was performed via variations of two general methods (*91*). One method is replacing water molecules with a step-wise increase of dehydration agents, allowing the crystal to equilibrate at each step. Dehydration agents used for Riok3 were glycerol, ethylene glycol and PEG-3350. Crystals of Riok3-Short were looped and transferred to mother liquor (original crystal condition) plus 5% dehydration agent and allowed to sit for 5 minutes. Once equilibrated, the crystal was looped again and transferred to mother liquor plus 10% dehydration agent. These 5% dehydration agent jumps and incubation steps were continued until dehydration agent reached mother liquor plus 30%. One crystal was allowed to rest overnight in the final 30% for each dehydration agent. This method of dehydration requires very robust crystals that can withstand a large amount of manipulation. RioK3-Short crystals did not appear to break down or become damaged during transfers. The other method of dehydration involves a single step allowing air to dehydrate crystals by evaporation of water molecules in a

reservoir of mother liquor containing the appropriate cryo condition and/or increased precipitant. RioK3-Short crystals were exposed to air for thirty minutes or one hour in 100 μL reservoirs containing mother liquor plus 20% ethylene glycol, mother liquor plus 10% PEG-3350, or mother liquor plus 10% ethylene glycol and 10% PEG-3350. Unfortunately, the diffraction of crystals from either method of dehydration did not improve diffraction.

All annealing and dehydration experiments were performed on the University of Maryland's Bruker home source macromolecular diffraction system. While the synchrotron is ideal to see small changes in diffraction quality, the home source was sufficient to determine if annealing or dehydration significantly improved diffraction, which did not occur.

Finally, the last type of crystal modification performed on RioK3-Short was seeding. There were two types of seeding attempted, macroseeding and microseeding (*91, 93, 94*). The names are very telling, as macroseeding is the process of looping a single large crystal, washing the crystal, and then placing the crystal for further growth in either an already equilibrated crystal condition or unequilibrated condition with fresh protein. Microseeding is pulverizing a small crystal into microscopic crystals, and then using a horse or cat whisker to transfer the microcrystals, or diluting the stock seed mixture, into an already equilibrated crystal condition below the precipitant concentration required for crystallization. The goal of microseeding is to set up a range of conditions with a decreasing number of nucleation points (microseeds) by streaking across multiple drops much like sequentially streaking bacteria onto an agar plate, with the purpose of acquiring a single colony. RioK3-

Short was subjected to both types of seeding. Microseeding did not yield any crystals, however macroseeding was successful in growing larger crystals. The larger crystals suffered from the same poor diffraction as the original pre-seeded crystals, with diffraction limited to ~6.0 Å when screened at the NECAT beamline.

7.3: RioK3-S199A-Short

Crystallography techniques to improve existing crystal diffraction were unsuccessful. Because RioK3-Short was soluble and a successful purification protocol was already determined, the next logical direction was the optimization of the RioK3-Short construct via site-directed mutagenesis. Pure RioK3-Short protein was sent previously for molecular weight determination via mass spectrometry at the University of Maryland Proteomics Core Facility managed by Dr. Yan Wang. The expected molecular weight of RioK3-Short (193-519) is 37,717.2 Da. The resulting spectra confirmed the expected MW of RioK3-short with a peak centered at 37,717.6 Da (Fig. 7.3A). A second peak centered at 37,794.7 Da suggested the presence of a singly phosphorylated version of RioK3-short because the difference between the two masses is the expected change for a post-translational phosphorylation modification. Excess protein was digested with trypsin and analyzed by tandem LC MS/MS by Dr. Yan Wang. Peptide determination of the trypsin fragments identified 84% of the sequence with only one possible site for phosphorylation (Fig. 7.3B). The phosphorylated residue was serine 199. Stratagene's Quikchange Lightning Site-Directed Kit was used to mutate serine 199 to alanine to abolish RioK3's ability to autophosphorylate. RioK3-S199A-Short purified identically to RioK3-Short with no deviations. Again, the S199A mutant was screened and surprisingly the same

condition in which RioK3-Short crystallized did not grow S199A crystals. Only one condition of 1152, containing 24% PEG-8000, 0.1 M Tris pH 8.0 and 0.2 M $MgCl_2$, grew crystals (Fig 7.4A). These crystals were optimized in the same manner as RioK3-Short and grew much larger (~50x500 μm) than anything previously acquired from RioK3-Short (Fig 7.4B). Even though the crystals appeared more ordered than the RioK3-Short crystals, the diffraction did not match that same visual improvement. In fact, diffraction images collected at the synchrotron were poorer and provided data to ~8.0 Å.

7.4: Surface Entropy Mutations

Surface entropy reduction (SER) is an accepted practice for improving crystallization chances for hard to crystallize proteins (*95, 96*). After the failure to improve diffraction past 5.5 Å over all previous forms of optimization, the RioK3 project needed a new crystal form, and SER could provide the change necessary to promote crystallization in a new form. SER is achieved by mutagenesis of surface residues with high conformational flexibility and thus a high amount of surface disorder/entropy. Mutation of these residues, which are typically lysine or glutamate, to alanine reduces this flexibility and reduces the entropy on the surface of the protein. The less disorder on the surface of the protein promotes tighter packing during crystallization, which leads to better diffracting, higher resolution crystals. It was shown that sequential mutations of residues directly next to each other improves the likelihood of crystallization and an online server (SERp) from UCLA can identify good targets for mutagenesis (*95, 97*). For RioK3, SERp identified two locations, each a series of three residues that are most likely to improve crystallization.

Residues 222-224 and 276-278 had the sequences Glu-Lys-Lys (EKK) and Glu-Lys-Glu (EKE), respectively. To make these mutations, Qiagen's Quikchange™ Lightning kit was used with RioK3-Short as the template DNA. For both locations, the mutants were first double mutated yielding constructs RioK3-KK223AA and RioK3-KE277AA, and then the third mutation was introduced creating RioK3-EKK222AAA and RioK3-EKE276AAA. For all four mutated constructs, there was no difference in the expression and successful purification protocol from the original RioK3-Short fragment. All four mutants were concentrated with no issue and subjected to crystallization screening via the Phoenix™. There were no crystals observed for RioK3-KE277AA and RioK3-EKE276AAA. Crystals were observed for RioK3-KK223AA and RioK3-EKK222AAA in the same condition and same crystal form as the original RioK3-Short. The inability to acquire a new crystal form from SER was disappointing and suggested an entirely new direction had to be taken to acquire a new crystal form for RioK3.

7.5: Limited Proteolysis

RioK3-Short, RioK3-S199A-Short, and all the surface entropy mutants were unsuccessful. The optimization methods described above exhausted crystallization trials for the 193-519 constructs. Focus shifted to determine a new RioK3 fragment that was both soluble and purifiable. The RioK3 full-length construct was expressed with a histidine tag in *E.coli* overnight and purified by affinity chromatography. Originally, the RioK3 full-length construct was insoluble. However, lessons learned from RioK3-Short (addition of ATP and $MgCl_2$) yielded small amounts of near pure soluble protein (~1 mg/L culture). The quantity of RioK3 full-length was not suitable

for crystallization trials, but fitting for limited proteolysis using thermolysin, trypsin, and elastase. The goal of the assay was to determine a stable construct that could be used for further crystallization trials. The time based incubation assay was designed to identify a gel band that was consistent over more than one time point and seen in more than one protease. For trypsin and thermolysin, 50 μL of Riok3 full-length concentrated to 1 mg/mL was mixed with 9 μL of enzyme buffer containing 100 mM Tris pH 8.0, 100 mM NaCl, 5 mM $CaCl_2$ and 1 μL of protease. For elastase, the enzyme buffer only contained 200 mM Tris pH 8.0 and was set up in the same volumes. The reaction was incubated at both 37 °C and on the bench at ~25 °C, with 15 μL samples taken at 5, 10, 30, and 120 minutes. The proteolysis reaction was stopped through denaturation by the addition of gel loading buffer containing SDS and heated to 95 °C. SDS-PAGE was used to analyze the samples (Fig 7.5A, B). Lane 3 from the elastase gel in Figure 7.5A shows a slice cut from the gel, highlighted by an arrow. This slice was sent to Dr. Yan Wang for complete in-gel tryptic digest and peptide determination via LC MS/MS. Dr. Wang was able to identify coverage from the gel band for residues 248-452 with 67% sequence coverage over that range. This result suggested a new design for the short constructs with deletions from both the N-terminus and the C-terminus. When combined with additional information from previously reported Rio1 archaeal structure and PSI-PRED secondary structure prediction tool, three new constructs (248-519, 248-474, and 193-474) were cloned with primer sets 5'-GAGAACCTGTACTTCCAGGGAATGTTGGAGACAATCACTGGC-3' and 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTTATCATTCATCATATAGTAGT

GGTGGGTC for 248-519, 5'-GAGAACCTGTACTTCCAGGGAATGTTGGAGAC AATCACTGGC-3' and 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTTATCA TGTGATGTTTAAGCCTGAAACAGC-3' for 248-474, and 5'-GAGAACCTGTACT TCCAGGGAATGGATTTAAAACTATCAAAC-3' and 5'-GGGGACCACTTTGTA CAAGAAAGCTGGGTTATCATGTGATGTTTAAGCCTGAAACAGC-3' for 193-474 utilizing previously described Invitrogen's Gateway™ cloning (Fig 7.5C) (*35, 98, 99*). These three constructs were subjected to expression and solubility testing. RioK3-248-474 and RioK3-248-519 were completely insoluble and could be collected in the cell pellet. RioK3-193-474 expressed and was partially soluble for the first step of purification. However, upon elution from the affinity column, the soluble protein immediately precipitated. The buffers for RioK3-193-474 were modified to prevent precipitation by varying the components (200 mM and 500 mM NaCl, ±1 mM ATP, ±10% glycerol). Buffer variations were unable to overcome RioK3-193-474's insolubility.

7.6: RioK3 Conclusions

My goal of solving the structure of RioK3 was an exhaustive lesson on the difficulties of both crystallography and protein purification. The only way to learn the "tricks" of crystallography, whether it be surface entropy mutations or an outside the box "X" method of optimization, is to gain experience through a challenging project. Ideally, one of the methods employed to improve diffraction works. However, this was not the case for RioK3. I confidently leave this project knowing the next person can successfully take lessons learned from my trials and apply that knowledge to acquiring a structure. I leave the RioK3 project with multiple soluble

constructs and a roadmap toward success based on the results from past experiments presented in this manuscript. My successes with expression, purification and crystallization overshadow the lack of a "hero" crystal. The next researcher fortunate enough to take over this project will be one step away from success. Maybe that researcher will continue to eliminate possible paths as I have during my tenure, or maybe that researcher will choose correctly on their first try. In any case, I believe my work will lead to success in the near future.

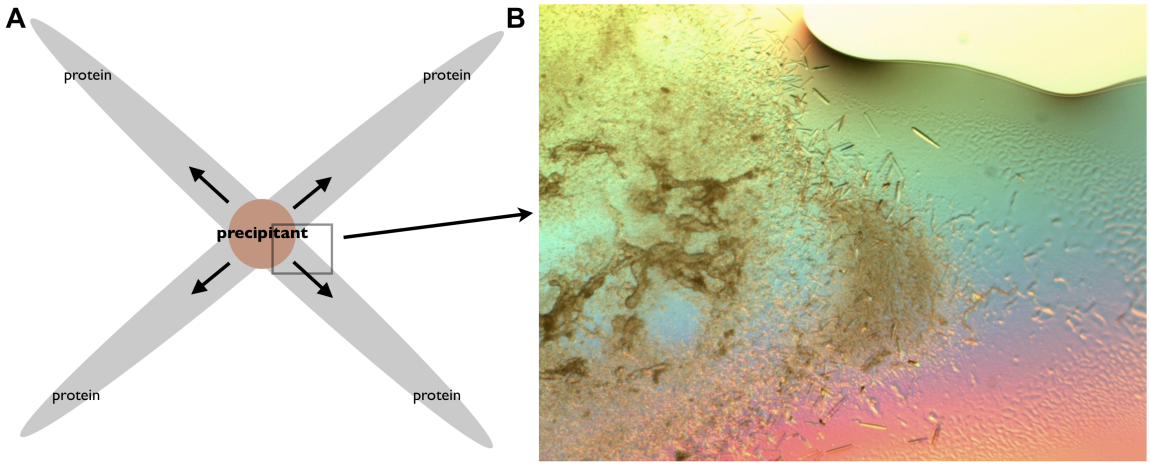


Figure 7.1: “X” Method

A) Cartoon representation of “X” method. Protein is streaked into the formation of an “X” with pipette tip and then precipitant is added to the center of the “X” as shown. **B)** Image taken with digital camera of crystals growing in the metaphase region created by dilution between precipitant near the center and soluble Riok3-Short on the corners. The box in (**A**) represents the location of the image taken for (**B**).

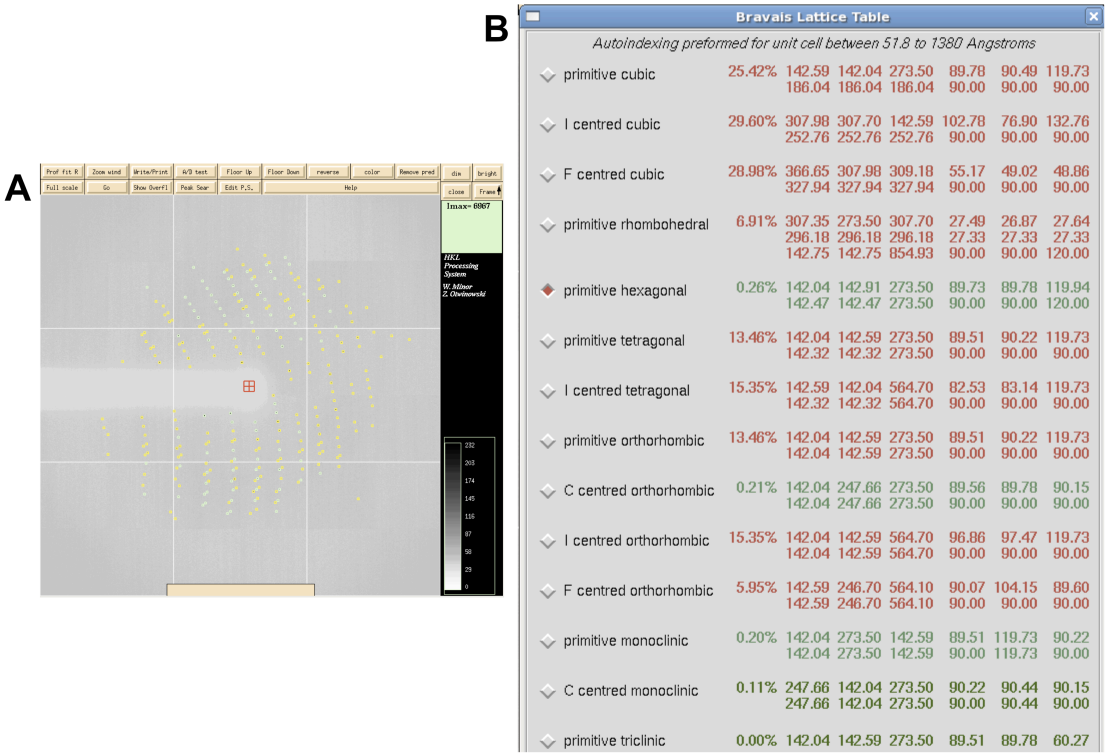


Figure 7.2: Indexing of Riok3-Short

A) Diffraction image from Riok3-Short crystal grown in minimal media showing the fitment of spots after refining in the primitive hexagonal point group. **B)** Choices of point groups determined form Riok3-Short diffraction image in (**A**). The highest symmetry with the lowest score (primitive hexagonal colored green) was chosen and confirmed during scaling, using Scalepack. DENZO, a program within the HKL2000 suite, was used for indexing (*56*).

Table 7.1: Riok3-Short Logfile

Shell		I/Sigma in resolution shells:								
Lower	Upper	% of reflections with I/Sigma less than								
limit	limit	0	1	2	3	5	10	20	>20	total
25.00	15.46	0.4	0.7	1.5	1.8	4.6	75.8	93.0	0.0	93.0
15.46	12.80	0.4	0.8	1.9	2.6	7.0	60.7	96.5	0.0	96.5
12.80	11.35	0.4	1.4	2.3	3.7	8.5	64.7	95.8	0.0	95.8
11.35	10.39	0.4	0.8	2.4	3.8	8.5	37.8	96.2	0.0	96.2
10.39	9.70	0.7	1.6	3.6	4.5	11.1	46.6	97.4	0.0	97.4
9.70	9.15	0.7	2.7	5.8	8.4	16.2	46.8	96.6	0.5	97.2
9.15	8.72	0.6	2.3	6.7	9.8	18.2	52.0	97.9	0.1	98.1
8.72	8.35	1.4	4.6	8.5	12.9	25.3	58.1	96.6	1.6	98.2
8.35	8.04	3.1	8.7	16.8	23.4	38.2	62.3	94.5	3.2	97.8
8.04	7.77	3.2	9.9	18.9	27.9	43.9	74.5	95.3	3.0	98.3
7.77	7.54	5.3	14.1	30.0	41.8	58.2	82.6	96.2	2.3	98.5
7.54	7.33	7.0	20.1	35.3	48.6	67.7	86.9	97.2	1.3	98.5
7.33	7.14	7.8	19.7	38.6	53.5	73.5	89.8	96.9	1.5	98.4
7.14	6.97	10.8	28.9	46.8	61.3	76.5	92.1	98.3	0.7	99.0
6.97	6.81	11.7	27.7	45.1	59.1	77.5	93.1	99.0	0.4	99.4
6.81	6.67	12.6	29.2	46.4	62.5	78.7	93.0	98.5	0.4	98.9
6.67	6.54	13.7	33.8	54.6	68.1	81.8	94.3	98.9	0.3	99.2
6.54	6.42	13.0	34.3	56.8	70.0	85.3	96.3	98.9	0.0	98.9
6.42	6.31	14.3	35.3	58.7	71.4	85.7	94.4	97.0	0.0	97.0
6.31	6.20	14.9	39.4	64.8	76.9	87.6	95.4	97.1	0.3	97.4
All	hkl	6.1	15.8	27.3	35.6	47.7	74.9	96.9	0.8	97.7

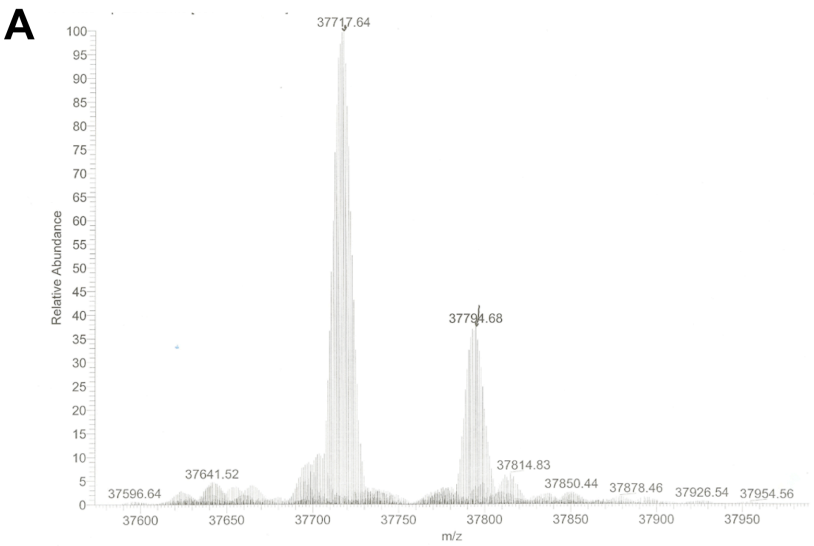
Shell			Average		Norm.	Linear	Square
Lower Limit	Upper Limit	I	error	stat.	Chi**2	R-fac	R-fac
15.46	12.80	203.0	21.6	8.8	1.350	0.098	0.115
12.80	11.35	210.7	23.6	8.6	1.431	0.104	0.122
11.35	10.39	203.7	18.1	8.3	1.665	0.093	0.102
10.39	9.70	157.9	14.9	8.3	1.560	0.103	0.112
9.70	9.15	115.7	10.5	6.9	1.730	0.108	0.113
9.15	8.72	111.2	10.5	7.1	1.691	0.111	0.112
8.72	8.35	93.9	9.2	6.8	1.525	0.114	0.118
8.35	8.04	62.9	7.4	6.4	1.386	0.139	0.133
8.04	7.77	43.7	6.8	6.3	1.254	0.171	0.179
7.77	7.54	32.7	6.6	6.3	1.198	0.225	0.214
7.54	7.33	25.3	6.7	6.4	0.942	0.252	0.242
7.33	7.14	22.6	6.9	6.7	0.946	0.296	0.280
7.14	6.97	20.1	7.1	6.9	0.829	0.318	0.274
6.97	6.81	19.8	7.3	7.1	0.885	0.370	0.357
6.81	6.67	18.6	7.5	7.2	0.964	0.384	0.354
6.67	6.54	16.6	7.8	7.5	0.825	0.421	0.376
6.54	6.42	15.3	8.0	7.7	0.846	0.485	0.437
6.42	6.31	14.0	8.2	7.9	0.924	0.488	0.454
6.31	6.20	13.0	8.6	8.3	0.829	0.493	0.431
All	reflections	79.7	11.1	7.3	1.234	0.133	0.120

Table 7.2: Matthews Coefficient Logfile

Matthews Coefficient logfile showing the number of possible monomers in the asymmetric unit, based on MW 37,717 Da. Highlighted in yellow are the most likely number of monomers.

For estimated molecular weight 37717.

Nmol/asym	Matt. Coeff	%solvent	P(6.18)	P(tot)
1	44.44	97.23	0.00	0.00
2	22.22	94.47	0.00	0.00
3	14.81	91.70	0.00	0.00
4	11.11	88.94	0.00	0.00
5	8.89	86.17	0.00	0.00
6	7.41	83.40	0.00	0.00
7	6.35	80.64	0.00	0.00
8	5.56	77.87	0.00	0.00
9	4.94	75.11	0.00	0.00
10	4.44	72.34	0.00	0.00
11	4.04	69.57	0.01	0.01
12	3.70	66.81	0.02	0.01
13	3.42	64.04	0.03	0.02
14	3.17	61.28	0.05	0.04
15	2.96	58.51	0.07	0.06
16	2.78	55.74	0.10	0.08
17	2.61	52.98	0.12	0.11
18	2.47	50.21	0.14	0.13
19	2.34	47.45	0.14	0.15
20	2.22	44.68	0.12	0.14
21	2.12	41.91	0.09	0.11
22	2.02	39.15	0.05	0.07
23	1.93	36.38	0.03	0.04
24	1.85	33.62	0.01	0.02
25	1.78	30.85	0.00	0.01
26	1.71	28.08	0.00	0.00
27	1.65	25.32	0.00	0.00
28	1.59	22.55	0.00	0.00
29	1.53	19.79	0.00	0.00
30	1.48	17.02	0.00	0.00
31	1.43	14.25	0.00	0.00
32	1.39	11.49	0.00	0.00
33	1.35	8.72	0.00	0.00
34	1.31	5.96	0.00	0.00
35	1.27	3.19	0.00	0.00
36	1.23	0.42	0.00	0.00



B

```
181        191        201        211        221        231
 |          |          |          |          |          |
181            GMDLKISN HVFNALKQHA YSEERRSARL HEKKEHSTAE KAVDPKTRLL   240
241 MYKMVNSGML ETITGCISTG KESVVFHAYG GSMEDEKEDS KVIPTECAIK VFKTTLNEFK   300
301 NRDKYIKDDF RFKDRFSKLN PRKIIRMWAE KEMHNLARMQ RAGIPCPTVV LLKKHILVMS   360
361 FIGHDQVPAP KLKEVKLNSE EMKEAYYQTL HLMRQLYHEC TLVHADLSEY NMLWHAGKVW   420
421 LIDVSQSVEP THPHGLEFLF RDCRNVSQFF QKGGVKEALS ERELFNAVSG LNITADNEAD   480
481 FLAEIEALEK MNEDHVQKNG RKAASFLKDD GDPPLLYDE
```

Figure 7.3: RioK3-Short MW Determination and LC MS/MS

A) Molecular weight determination of RioK3-Short. The expected MW for RioK3-Short is 37,717 Da. **B)** Results from peptide identification via tryptic digestion followed by LC MS/MS. Red residues indicate identified residues with a box around serine 199, the only phosphorylated residue identified. Dr. Yan Wang at the University of Maryland Proteomics Center performed all mass spectrometry for RioK3. Data was analyzed using Mascot (www.matrixscience.com).

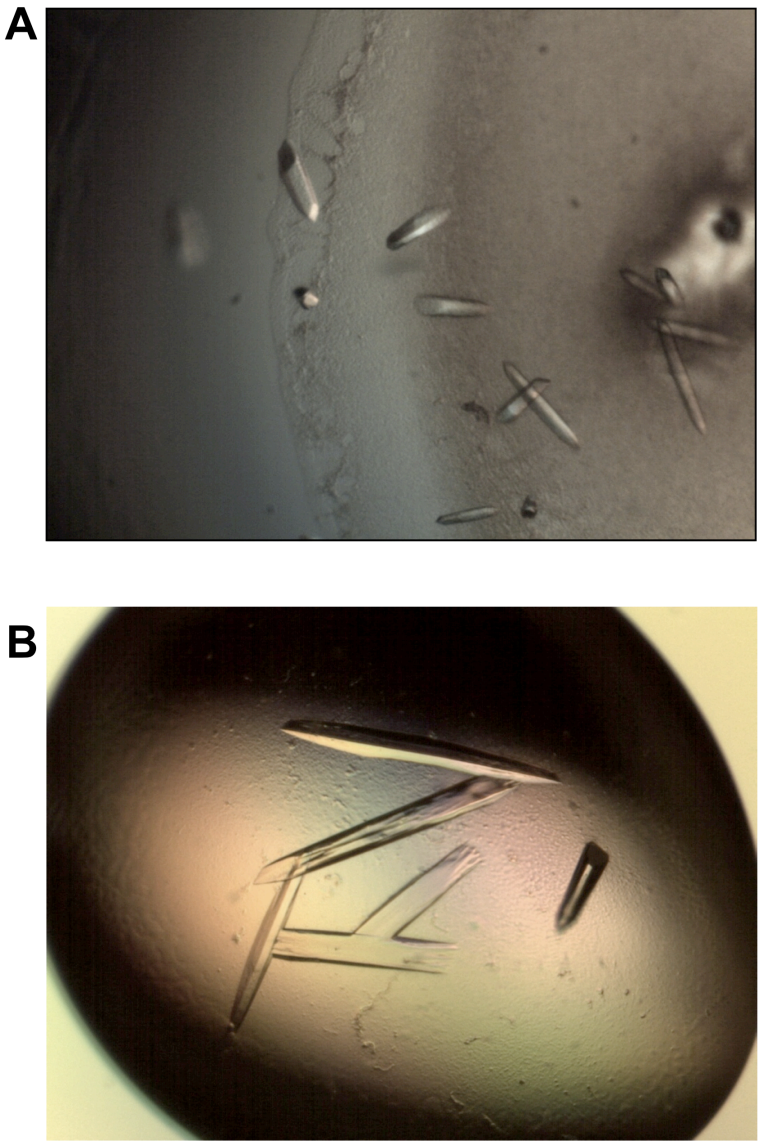


Figure 7.4: RioK3-S199A-Short Crystals

A) Digital image of RioK3-S199A-Short crystals grown in the Wizard II screen position E3, containing 24% PEG-8000, 0.1 M Tris pH 8.0, 0.2 M $MgCl_2$. **B)** The largest crystals grown for RioK3. This image was photographed from an optimization tray containing 20% PEG-8000, 0.1 M Tris pH 8.5, 0.2 M $MgCl_2$.

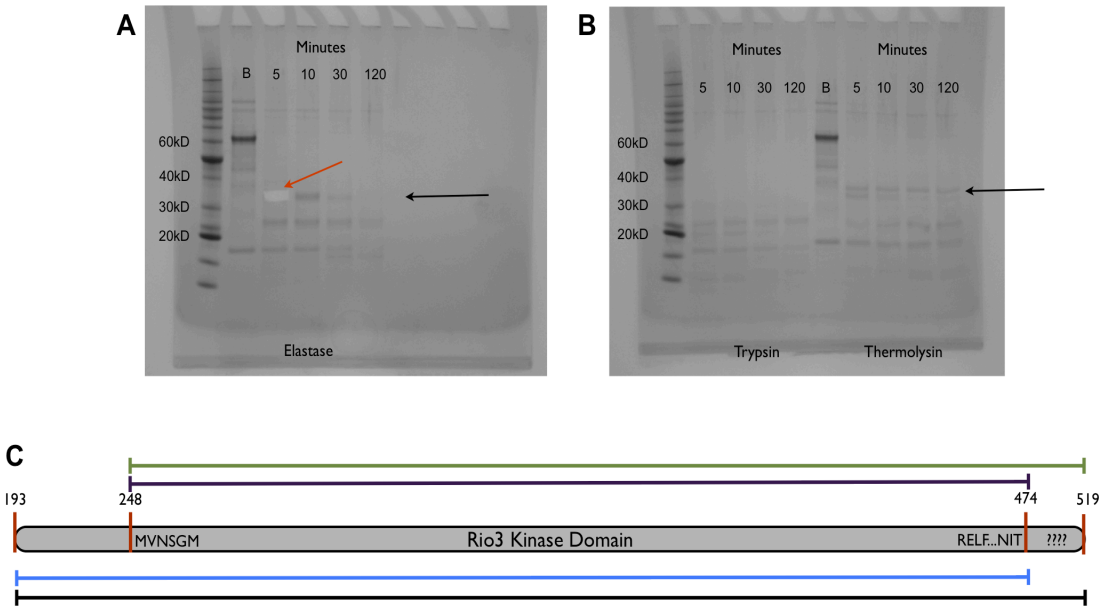


Figure 7.5: Limited Proteolysis of Riok3 Full-length

SDS-PAGE of RioK3 full-length subjected to limited proteolysis with elastase, shown in (**A**), trypsin or thermolysin, both shown in (**B**). Horizontal arrows point to band of interest and the red diagonal arrow in (**A**) points to gel slice removed for mass spectrometry. **C)** Representation of fragments created from limited proteolysis assay (green = 248-519, purple = 248-474, blue = 194-474, black = original Riok3-Short 193-519). MVNSGM cuts off the "extended" archaeal Rio1 domain, supported by the MS/MS results (*35*). RELF-NIT is homologous to the last archaeal Rio1 alpha helix I (*35*). "????" represent unfolded region predicted from PSI-PRED (*99*).

Chapter 8: My Role in Ubiquitin

8.1 Overview and Thanks

I would like to thank Dr. David Fushman at the University of Maryland, a member of the Center for Biomolecular Structure and Organization (CBSO), for the opportunity to learn and teach while working on his Ubiquitin projects. My role in Ubiquitin has been two fold. First and foremost, my responsibility was to provide advice and consultation to members of his laboratory who had an interest in learning the basics of crystallography. Shirley Lee, current graduate student at the University of Michigan was the primary contact for Dr. Fushman. My second purpose for the multiple Ubiquitin projects was to provide expertise and resources not available to Dr. Fushman. This goal was met by providing beamtime during synchrotron trips and access to instrumentation in the Dr. LaRonde-LeBlanc lab, along with guidance on all shared instrumentation. It should be noted that my role in Ubiquitin was strictly crystallographic in nature, meaning I only provided suggestions and guidance in crystal growth/optimization, crystal harvesting, data collection, and structure solution. I was not responsible for project direction or the selection of crystallographic targets. Many of the targets chosen by Dr. Fushman were site-directed mutants of human Ubiquitin, designed to look at confirmation changes introduced by the selected mutations. Three of these mutants provided crystals that diffracted to 2.0 Å or better. In chronological order, Ubiquitin leucine 69 to serine (L69S), isoleucine 13 to serine (I13S), and double mutant lysine 63 to aspartatic acid with glutamatic acid 64 to glycine (KE63DG) were all selected for x-ray structure determination. Near the end

of my tenure at the University of Maryland, I also helped with the structure determination of lysine 48-linked di-ubiquitin in complex with E2-25K, the protein known to be responsible for creating chains of ubiquitin through a covalent bond between lysine 48 and glycine 76 (*100*). The following is a description of results from my time spent with Dr. Fushman and his laboratory members.

8.2: Methods for Structure Determination of Ubiquitin Mutant I13S

Initial crystallization was done via sparse matrix screening (Wizard I & II & III™, Natrix™, PEG™ Suite, Cryo™ Suite, Index™) using Art Robbins Phoenix™ high-throughput robot. The Phoenix utilizes the sitting drop method for screening for crystallization conditions. Ubiquitin I13S mutant crystals grew in 5-7 days. Optimization of crystals screened for diffraction were done via hanging drop method using a 24-well tray purchased from Qiagen. Crystallization conditions for Ubiquitin I13S mutant include 0.3M Ammonium Iodide, 26% PEG 3350, 100mM Hepes pH 6.8, and 20% glycerol. Crystals were screened for diffraction and cryo conditions at the home souce (Microstar H2 generator with Proteum Pt135 CCD detector) at 100 K. Diffraction data sets were collected via Advanced Photon Source (APS) located at Argonne National Laboratory NECAT Beamline 24IID. Ubiquitin crystal datasets were collected at a wavelength set at 1.00931 Å. All data collected was integrated and scaled using HKL2000 (Table 8.1) (*56*). The human Ubiquitin monomer (PDB: 1ubq) was used as the search model for molecular replacement using Phaser within CCP4, for the structure solution of Ubiquitin I13S mutant (*66, 70, 101*). Refinement of the model was done primarily using Refmac5 within CCP4 and model building was done using COOT (*64-66*). The final solution intended for submission was

refined with PHENIX (*72*). I13S crystallized with four molecules in the asymmetric unit in space group $P2_1$. Figure 8.1 shows an overall cartoon view of the asymmetric unit. Statistics for I13S data collection and refinement are shown in Table 8.2.

8.3: Ubiquitin Mutants L69S and KE63DG

Crystals were obtained for L69S and KE63DG in the same manner as I13S (screening then optimization) and data was collected at both the home source and synchrotron. Unfortunately, both L69S and KE63DG were not solved due to problems with data quality. Specifically, both crystal forms used for data collection suffered from pseudosymmetry-related twinning that could not be resolved. Pseudosymmetry confuses space group selection due to non-crystallographic symmetry being close to crystallographic symmetry. This results in two or more possible spacegroups contained within the crystals. In this case, we were unable to successfully de-twin the data and molecular replacement could not effectively place molecules to provide accurate phases. This resulted in poor electron density and incorrect solutions.

The first approach to resolving pseudosymmetry is going back to the original data to determine if there is a possibility that this space group confusion is caused by overlapping crystals, otherwise known as non-merohedral twins (common in plate crystals), thus giving two distinct overlaid diffraction patterns. After taking a closer look at diffraction patterns and scaling statistics, multiple crystals did not appear to be the cause of twinning. The presence of a single lattice in the diffraction pattern and confusion between symmetry suggested pseudo-merohedral twinning. Pseudo-merohedral twinning is the presence of multiple lattices of differing symmetries

directly overlaid on top of each other giving the appearance of a single crystal lattice. The next approach in this type of twinning was to re-index the collected data into lower symmetry, in this case primitive monoclinic (P2) instead of primitive orthorhombic (P222). Initially, the most likely space group determined for both L69S and KE63DG was $P2_12_12_1$ (orthorhombic), which can be confused with $P2_1$ (monoclinic) in a case of pseudosymmetry. Ideally, the elimination of symmetry not actually present should result in a usable dataset. This was not the case for the Ubiquitin mutants evident by the continued failure of molecular replacement. This suggested a more severe partial twinning that would require separation of the twinned data using a partial twin law. I was never successful in finding an appropriate method for solving these mutant structures. In hindsight, work should have focused on the growth of many crystals and required full-time dedication to optimization of these twinned Ubiquitin crystals.

8.4: E2-25K in Complex with Di-ubiquitin

Near the end of my time spent with Shirley Lee, she grasped many of my teachings and was fully capable of making decisions and providing assistance to other members of Dr. Fushman's laboratory. Ms. Lee provided her knowledge to help Dr. Rajesh Singh, postdoctoral reseacher in Dr. Fushman's lab, prepare E2-25K and lysine 48 linked di-Ubiquitin protein complex and set up screening trays using the Phoenix™ liquid handler. Shortly after setting up the initial screening, Ms. Lee left the University of Maryland and beginning graduate school at the University of Michigan. I picked up where Ms. Lee left off and began work with Dr. Singh. The results from the screen were very encouraging with over 50 possible targets for

optimization. There were two distinct crystals forms, clean plates and circular "blob" crystals (Fig 8.2A, B). I utilized the home diffraction source to collect a data set from both plate crystals and a "blob" crystal and confirmed the presence of the complex only in the "blob" form via a low-resolution (3.0 Å) structure solved by molecular replacement (*72*). This structure in $P2_12_12_1$, solved with E2-25K and one monomer of Ubiquitin in the asymmetric unit. Crystallographic symmetry was present between the two monomers of linked Ubiquitin, resulting in the single monomer of Ubiquitin. There was no discernible electron density to model residues 73-76 between the Ubiquitin monomer and its symmetry related mate, and the covalent linkage between lysine 48 and glycine 76. This could have been due to low-resolution diffraction, the crystallographic axis between the two monomers, or flexibility in the short polypeptide connecting the Ubiquitin monomers. The synchrotron was needed to collect higher resolution data. Crystals harvested from Qiagen's Cryos™ Suite containing 10.2% PEG-20000, 0.085 M MES pH 6.5 and 15% glycerol, diffracted to 2.3Å at the NECAT beamline (Table 8.3). Molecular replacement was performed with Phaser within PHENIX, using a previously solved crystal structure of E2-25K and ubiquitin complex (PDB: 3K9P) (*70, 72, 102*). In space group $P2_1$, there are two monomers of E2-25K and two monomers of Ubiquitin (Fig. 8.3A). Residues 74-76, including the covalent linkage between lysine 48 and glycine 76, still could not be modeled due to poor electron density insufficient for accurate placement of those residues (Fig 8.3B). This suggests high conformational flexibility for these residues. Statistics for E2-25K and Ubiquitin complex are shown in Table 8.2.

8.5: Ubiquitin Conclusions

Again, it has been an incredible opportunity to provide assistance to Ms. Lee, Dr. Singh, and Dr. Fushman. Ubiquitin has truly been a lesson in both the successes and pitfalls of crystallography.

Table 8.1: Ubiquitin I13s Scalepack Logfile

Shell Lower limit	Upper limit	Average Redundancy Per Shell
40.00	4.31	3.9
4.31	3.42	3.8
3.42	2.99	4.0
2.99	2.71	4.1
2.71	2.52	4.1
2.52	2.37	4.1
2.37	2.25	3.9
2.25	2.15	3.7
2.15	2.07	3.6
2.07	2.00	3.5
All	hkl	3.9

Shell		I/Sigma in resolution shells:								
Lower	Upper	% of reflections with I/Sigma less than								
limit	limit	0	1	2	3	5	10	20	>20	total
40.00	4.31	0.3	1.4	2.2	2.8	4.1	8.1	17.9	81.7	99.6
4.31	3.42	0.1	0.6	1.2	1.8	3.8	9.4	22.9	76.5	99.5
3.42	2.99	0.5	1.3	3.0	4.7	8.2	19.2	40.9	58.8	99.8
2.99	2.71	1.8	4.7	7.8	10.8	17.2	31.0	56.7	43.1	99.8
2.71	2.52	1.8	5.5	9.5	14.1	22.5	40.9	67.1	32.9	100.0
2.52	2.37	2.2	7.0	13.5	19.4	29.4	47.7	72.5	27.4	99.9
2.37	2.25	2.7	9.6	17.9	25.4	35.9	54.8	78.5	21.3	99.8
2.25	2.15	3.3	12.1	21.3	29.7	40.8	58.2	80.1	18.1	98.1
2.15	2.07	3.9	14.5	24.8	32.8	43.3	61.5	81.4	12.0	93.5
2.07	2.00	4.3	14.0	24.9	32.9	43.0	60.0	76.8	8.8	85.6
All	hkl	2.1	7.0	12.5	17.3	24.7	38.9	59.3	38.3	97.6

Shell			Average		Norm.	Linear	Square
Lower Limit	Upper Limit	I	error	stat.	Chi**2	R-fac	R-fac
40.00	4.31	1072.7	22.9	14.1	2.757	0.047	0.050
4.31	3.42	983.8	24.4	17.2	2.983	0.059	0.060
3.42	2.99	495.4	15.6	12.6	2.792	0.078	0.074
2.99	2.71	305.6	12.3	10.7	2.105	0.088	0.082
2.71	2.52	209.2	11.1	10.1	1.722	0.106	0.098
2.52	2.37	194.8	11.5	10.5	1.542	0.110	0.096
2.37	2.25	161.3	12.0	11.3	1.357	0.123	0.115
2.25	2.15	147.8	13.0	12.4	1.338	0.132	0.126
2.15	2.07	118.4	13.3	12.8	1.158	0.155	0.154
2.07	2.00	108.2	13.8	13.3	1.037	0.160	0.154
All	reflections	388.7	15.1	12.5	1.924	0.077	0.064

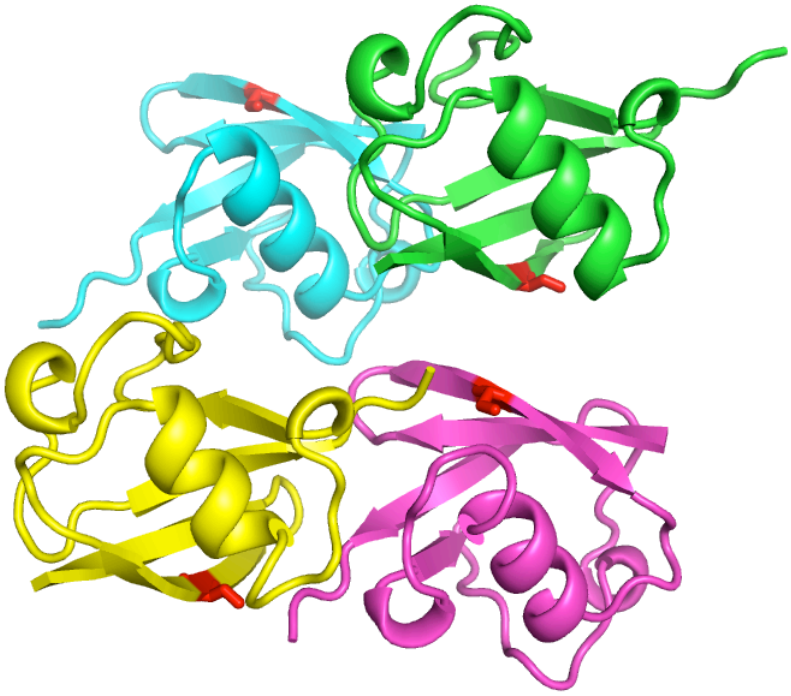


Figure 8.1: Ubiquitin I13S Overall View

Cartoon Representation of the four Ubiquitin monomers (colored cyan, green, yellow, and magenta) in the asymmetric unit. Serine 13 is mutated from isoleucine and shown highlighted in red sticks.

Table 8.2: Ubiquitin Data Collection and Refinement Statistics

	Ubiquitin I13S	**E2-Diubiquitin**
PDB code	**N/A**	**N/A**
Data collection		
Space group	$P2_1$	$P2_1$
Cell dimensions		
a, b, c (Å)	42.6, 58.1, 51.3	42.1, 38.5, 191.1
α, β, γ (°)	90.0, 91.0, 90.0	90.0, 90.0, 90.0
Molecules/Asym. unit	4	4
Wavelength (Å)	1.00931	0.97918
Resolution (Å)	40-2.0	50-2.3
R_{sym} (last shell)	0.077 (0.160)	0.074 (0.377)
$I/\sigma I$	25.7 (7.8)	22.6 (1.7)
Completeness (%)	97.6 (85.6)	96.4 (87.3)
Redundancy	3.9 (3.5)	3.5 (3.2)
Refinement		
Resolution (Å)	40-2.0	40-2.3
R_{work}/R_{free} (%)	17.5/22.9	18.4/23.4
No. non protein molecules		
Solvent	179	122
Mean B-factors (Å)^2 (TLS refined)	24.6	42.5
RMS deviations		
Bond lengths (Å)	0.007	0.008
Bond angles (°)	1.058	1.085
Ramachandran plot		
Most favorable	95.4%	92.0%
Additionally allowed	3.8%	7.4%
Generously allowed	0.0%	0.6%
Disallowed	0.8%	0.0%

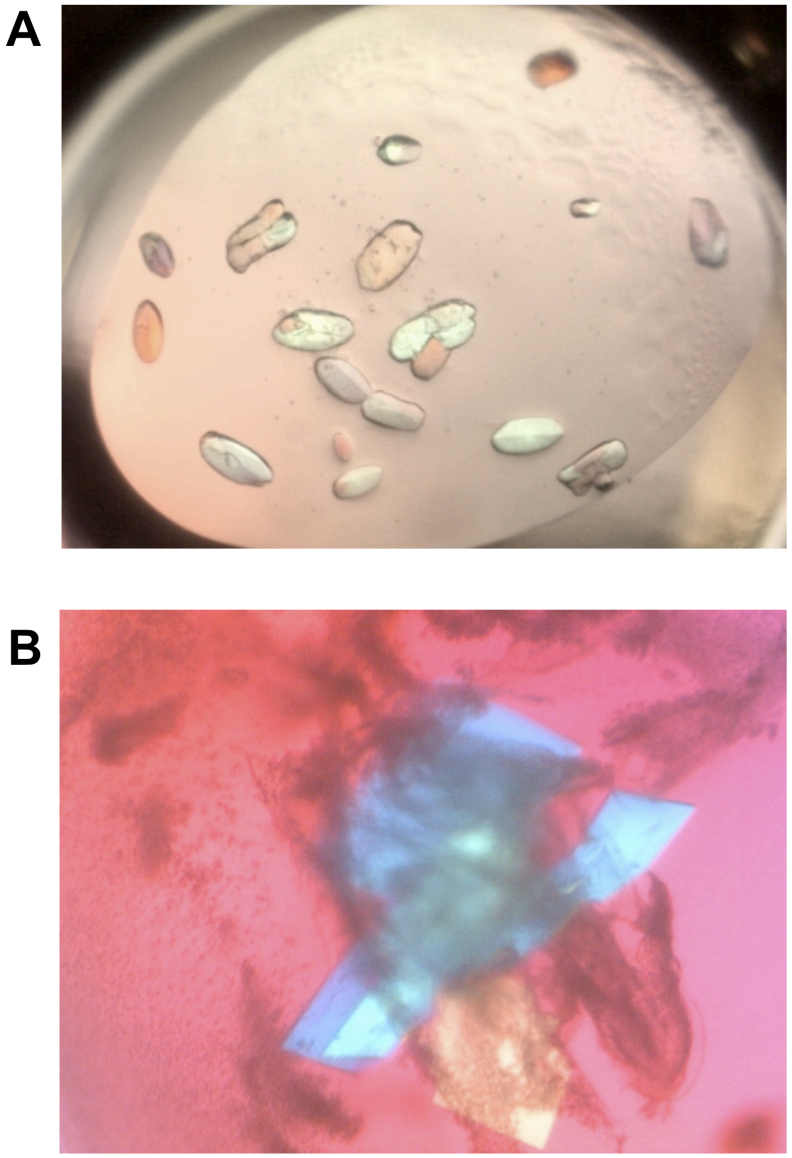


Figure 8.2: Possible Crystals Containing E2-25K and Di-Ubiquitin Complex

A) Digital image of one of many conditions that grew "blob" like crystal form. This crystal form was confirmed to contain E2-25K and di-Ubiquitin complex. **B)** Digital image of plate like crystals that also grew in many conditions (>30). This crystal form was later confirmed at the synchrotron to only contain Ubiquitin and not the E2-25K and di-Ubiquitin complex.

Table 8.3: E2-25K Diubiquitin Scalepack Logfile

Shell Lower limit	Upper limit	Average Redundancy Per Shell
50.00	4.95	3.5
4.95	3.93	3.6
3.93	3.44	3.5
3.44	3.12	3.5
3.12	2.90	3.5
2.90	2.73	3.5
2.73	2.59	3.5
2.59	2.48	3.5
2.48	2.38	3.4
2.38	2.30	3.2
All	hkl	3.5

Shell		I/Sigma in resolution shells:								
Lower	Upper	% of reflections with I/Sigma less than								
limit	limit	0	1	2	3	5	10	20	>20	total
50.00	4.95	1.1	2.2	3.5	4.2	5.3	9.8	23.7	76.0	99.7
4.95	3.93	0.9	2.1	3.3	5.0	7.4	16.1	75.6	24.3	100.0
3.93	3.44	1.2	3.2	5.3	7.5	11.8	23.3	63.7	35.7	99.4
3.44	3.12	2.0	5.7	9.8	13.9	22.6	40.9	72.3	26.7	99.0
3.12	2.90	3.8	9.4	17.0	23.0	34.1	54.5	80.1	18.2	98.3
2.90	2.73	4.9	14.4	25.5	34.2	46.8	70.1	90.0	7.6	97.5
2.73	2.59	6.5	19.3	33.9	44.5	59.9	79.1	92.9	3.6	96.6
2.59	2.48	7.8	22.7	38.0	50.3	66.9	84.8	93.5	1.6	95.1
2.48	2.38	9.8	28.5	45.1	57.4	71.7	85.1	90.6	0.5	91.1
2.38	2.30	11.1	31.5	50.3	62.9	74.3	84.4	87.1	0.2	87.3
All	hkl	4.9	13.8	23.0	30.0	39.7	54.4	76.6	19.9	96.4

Shell			Average		Norm.	Linear	Square
Lower Limit	Upper Limit	I	error	stat.	Chi**2	R-fac	R-fac
50.00	4.95	85.6	2.3	1.8	0.945	0.029	0.028
4.95	3.93	78.3	4.3	1.8	1.070	0.064	0.076
3.93	3.44	51.2	2.6	1.6	1.432	0.073	0.076
3.44	3.12	27.4	1.7	1.4	1.094	0.078	0.074
3.12	2.90	16.1	1.3	1.3	0.930	0.099	0.087
2.90	2.73	10.0	1.3	1.3	0.810	0.140	0.121
2.73	2.59	7.1	1.3	1.3	0.767	0.192	0.161
2.59	2.48	5.5	1.3	1.3	0.717	0.238	0.199
2.48	2.38	4.2	1.3	1.3	0.753	0.321	0.262
2.38	2.30	3.4	1.4	1.4	0.693	0.377	0.312
All	reflections	30.2	1.9	1.4	0.933	0.074	0.061

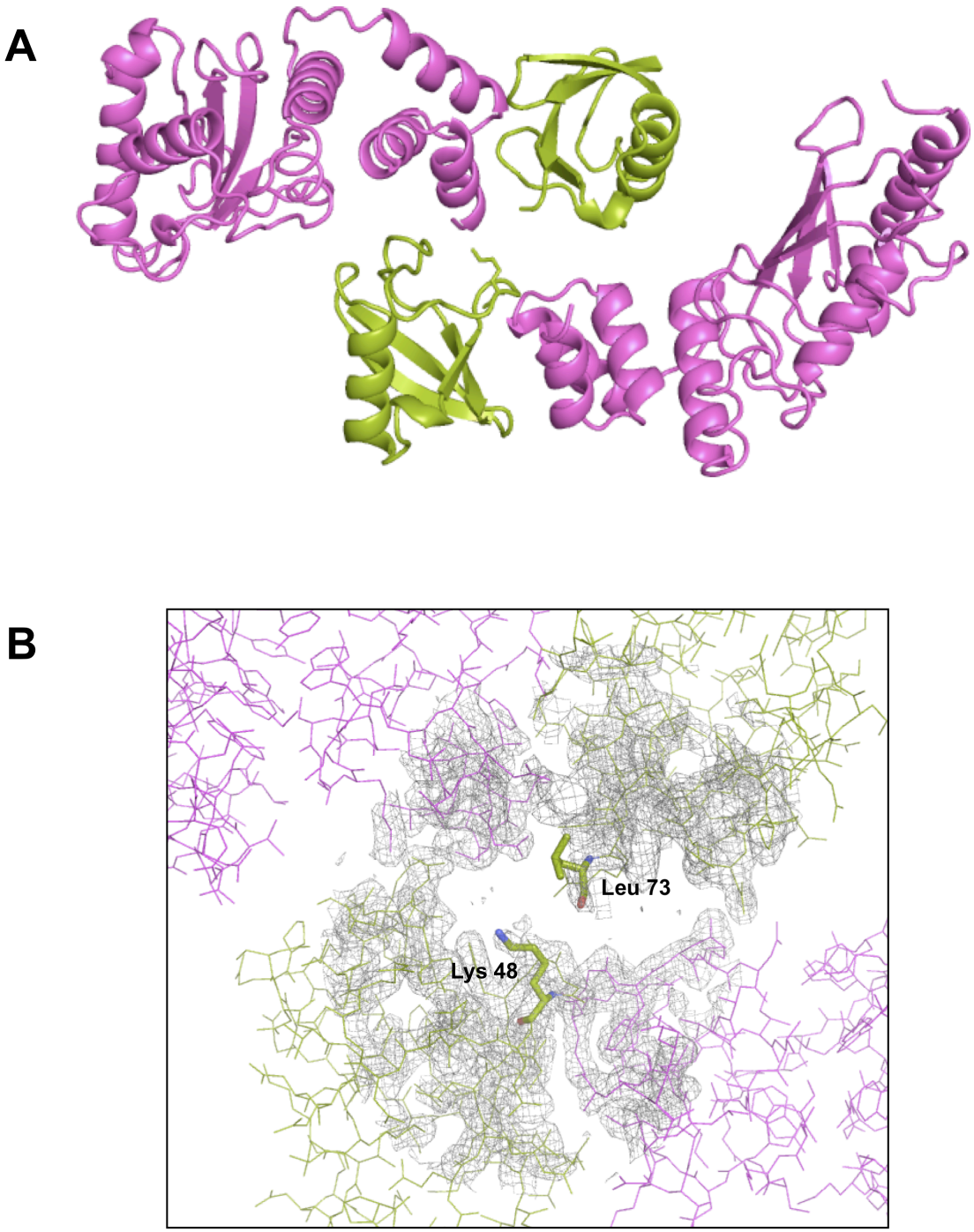


Figure 8.3: E2-25k and Di-Ubiquitin Complex

A) Cartoon representation of the contents of the E2-25K and di-Ubiquitin complex asymmetric unit, which includes two molecules of E2-25K (magenta) and two molecules of Ubiquitin (green). B) Electron density 2Fo–Fc map contoured at 1σ,

showing the gap in electron density between the last built residue (leucine 73) and the expected covalent bond with lysine 48.

Chapter 9: Crystal Structure of Iodotyrosine Deiodinase, a Novel Flavoprotein Responsible for Iodide Salvage in Thyroid Glands

9:1: Overview

Chapter 9 includes portions of the work published with Dr. Patrick McTamney, Jennifer Adler, and Dr. Steven Rokita. As first author on the article, and author of this dissertation manuscript, it is my commitment to discuss my responsibilities for the project and those responsibilities for which I was not the primary scientist. This chapter will include modified text from the article including crystallographic methods (crystal screening, data collection, structure determination and refinement). Included, and not present in the published article, are crystallographic tables (Tables 9.1, 9.2, 9.3-9.5) representing the original data used to solve the protein structures. Results and Discussion has been limited here, the full article can be found through the *Journal of Biological Chemistry* (*103*).

Dr. Patrick McTamney was responsible for all expression and purification of iodotyrosine deiodinase (IYD) protein. His work to achieve a soluble construct was exemplary and not described in this dissertation manuscript. The full experimental procedure for expression/purification work can be found through the University of Maryland library system in Dr. Patrick McTamney's recently published dissertation (2009, http://hdl.handle.net/1903/9848).

Jennifer Adler contributed to growing crystals and provided an excellent figure depicting the alignment of secondary structure between IYD and family members. This alignment figure is not included with from the following and can be

found online in the official published copy from the *Journal of Biological Chemistry* (*104*).

As senior author and the principle investigator for this project, Dr. Steven Rokita provided expert discussion and background knowledge for the IYD project. Dr. Rokita wrote all active site mechanistic results and discussion. Also Dr. Rokita provided an in-depth introduction, which was essential in providing the basis for structural analysis and guidance throughout the article.

Finally, I have to acknowledge my advisor Dr. Nicole LaRonde-LeBlanc. She drove the crystallization and structure determination part of this project to success and provided all the necessary resources, both in terms of expertise and instrumentation, to complete this work.

9.2: Statement of Authorization

This research was originally published in Journal of Biological Chemistry. Seth R. Thomas, Patrick M. McTamney, Jennifer M. Adler, Nicole LaRonde-LeBlanc, Steven E. Rokita. Title Crystal Structure of Iodotyrosine Deiodinase, a Novel Flavoprotein Responsible for Iodide Salvage in Thyroid Glands. *Journal of Biological Chemistry*. 2009; 284(29):19659-67. © the American Society for Biochemistry and Molecular Biology.

9.3: Introduction

Iodotyrosine deiodinase (IYD) salvages iodide from mono-iodotyrosine (MIT) and di-iodotyrosine (DIT) (*105, 106*). MIT and DIT are by-products from the proteolysis of thyroglobulin, a large protein (>300kD) responsible for the

biosynthesis of thyroxine (*105*). IYD is responsible for salvaging the iodide from MIT and DIT before excretion via a dehalogenation reaction for the purpose of recycling the iodide back into thyroglobulin, and subsequently the biosynthesis of thyroxine in the thyroid. If iodide is not recycled properly by IYD, then the acquisition of iodide from outside sources (diet) becomes the only pathway to gather iodide into the thyroid (*107, 108*). If for any reason the quantity of available iodide is insufficient for the biosynthesis of the essential thyroid hormones, then the unfortunate individual may suffer from hypothyroidism. Weight gain, fatigue, depression, weakness, constipation, and many other symptoms characterize hypothyroidism, resulting in a very serious condition (*109*). Currently, the most successful treatment for hypothyroidism is an oral supplement to replace insufficient levels of thyroxine (*109*). However, the treatment does not help everyone and more fundamental research needs to be performed to better understand the homeostasis of iodide in the human body. Presented in this chapter are three structures of IYD, one with IYD and cofactor flavin mononucleodite (FMN) and two structures solved in the presence of both cofactor and their substrate MIT or DIT, at 2.0 Å, 2.45 Å, and 2.6 Å, respectively. These structures provide an understanding for IYD mutations found in patients suffering from hypothyroidism.

9.4: Experimental Procedures

9.4A: Gene Construction of a Soluble and Affinity-tagged IYD

Gene construction of the soluble fragment used for crystallography is attributed to the hard work of Dr. Patrick Mctamney and Dr. Steven Rokita. The full

description of cloning can be accessed in the experimental procedures section of Thomas et al. (*103*) and in Dr. Patrick McTamney's dissertation (2009, http://hdl.handle.net/1902/9848). Briefly, the successful construct was truncated by 33 amino acids on the N-terminus to remove the membrane anchor domain and included an uncleavable His_6 tag on the C-terminus for purification. This truncation resulted in a soluble construct that met the purity and quantity requirements needed for crystallography.

9.4B: Expression and Purification of IYD

Dr. Patrick McTamney and Dr. Steven Rokita were responsible for the expression and purification of IYD. One of the major factors in determination of the feasibility of IYD structure determination was their ability to provide ample amount of pure protein for crystal screening and optimization. The full description of the expression of IYD in Sf9 cells and purification by affinity chromatography can be accessed in the experimental procedures section of Thomas et al. (*103*) and within Dr. Patrick McTamney's dissertation (2009, http://hdl.handle.ne/1902/9848).

9.4C: Crystallization

Initial crystallization was explored with sparse matrix screening (Wizard™ I, II, and III (Emerald Biosciences); PEGSuite™ and CryoSuite™ (Qiagen); and Natrix™ and Index™ (Hampton Research)) using an Art Robbins Phoenix™ high throughput liquid handler. Ultimately IYD was crystallized at 20 °C in three different crystal forms by the hanging drop vapor diffusion method using two parts IYD (16 mg/ml; 10 mM potassium phosphate, pH 7.4) to one part reservoir solution. IYD

crystals were obtained using a reservoir solution containing 20% (w/v) polyethylene glycol 3000, and 0.1 M sodium acetate, pH 4.5. IYD·MIT co-crystals were obtained by supplementing the enzyme solution with 2 mM MIT prior to addition of the reservoir solution containing 0.2 M magnesium chloride, 20% (w/v) polyethylene glycol 3350, 10 mM potassium phosphate, pH 7.4. IYD·DIT co-crystals were obtained by supplementing the enzyme solution with 2 mM DIT prior to addition of the reservoir solution containing 0.2 M ammonium acetate, 45% 2-methyl-2,4-pentanediol, 0.1 M BisTris, pH 5.5. In each case, yellow crystals appeared within 24 h. Crystals grew as connected clusters that required dehydration and micro-dissection to separate out single crystals.

9.4D: Data Collection

Initial crystals were screened for diffraction and cryo conditions using a Bruker Microstar H2 generator with Proteum Pt135 CCD detector at 100 K. Diffraction data sets were then collected at the Northeastern Collaborative Access Team Beamline 24-ID, Advanced Photon Source, Argonne National Laboratories. Diffraction data for IYD was collected at a wavelength of 1.653 Å for single anomalous diffraction phasing using sulfur atoms. In order to obtain the required redundancy and signal to noise ratio for the sulfur anomalous signal, data sets were collected using varying parameters including exposure time to minimize radiation damage. Diffraction data of co-crystals containing MIT and DIT alternatively were collected at 0.9795 Å. All collected data were integrated and scaled using HKL2000 (*56*).

9.4E: Sulfur Phasing and Refinement

All nine sulfur positions within IYD were found using SHELX within HKL2MAP program suite (Table 9.1) (*110-112*). Heavy atom positions were refined, and initial phases were determined using the AutoSHARP program set (*63*). Initial model building was performed by WARP within AUTOSHARP (*63*). Because of the complexity of the monomeric fold, nearly half of the residues were not built and required manual building. Several rounds of model building and refinement were needed for IYD using Refmac5 with TLS refinement and COOT for molecular visualization and rebuilding (*64-66*). A monomer of IYD was used as a model for molecular replacement to solve the structure of the co-crystal containing MIT. One dimer of this co-crystal was used in turn as the model for molecular replacement to solve the co-crystal containing DIT. Molecular replacement was performed by PHASER within the CCP4 program suite (*66, 70*). These structures were also rebuilt and refined by iterations of Refmac5 and COOT (*64*). IYD·DIT was initially solved in space group $P2_12_12_1$ and suffered from pseudosymmetry, evident after multiple rounds of refinement with high R-factors. Re-indexing IYD·DIT diffraction data into primitive monoclinic and subsequent determination of the structure in $P2_1$ removed the overestimated symmetry and allowed for proper refinement. Refinement statistics are provided in Table 9.2. Tables 9.3-9.5 show data collection log files for all three structures matching those submitted with Table 9.2.

9.5: Results

9.5A: Overall Structure of the Soluble Domain of IYD

In the absence of substrate, IYD crystallized with one monomer per asymmetric unit. However, analysis of the crystal packing interactions revealed an extensive interface with a symmetry-related molecule (2490 $Å^2$ per monomer) indicating that the enzyme forms a domain-swapped dimer as observed in related proteins within the same structural superfamily (*113-116*). The IYD·MIT complex crystallized with two monomers per asymmetric unit, and the IYD·DIT complex crystallized with eight monomers per asymmetric unit. All three structures were solved from distinct crystal forms (Table 9.2). Electron density for the first 34 amino acids of the truncated enzyme was not observed for any of the three crystals. These residues in the native protein likely provide a flexible linker between the soluble domain and the N-terminal anchor embedded in the membrane.

As mentioned, IYD contains the characteristic α-β fold that is common to all proteins of the NADH oxidase/flavin reductase superfamily (Fig. 9.1, Supplementary Fig. S1). The closest structural neighbor to IYD as determined by DaliLite (*117*) is BluB (Protein Data Bank code 2ISK) (*116*) with an (RMSD) of 3.1 Å for 198 structurally equivalent residues of 219 possible residues. Despite the structural similarities, their sequence identity is quite low (19%). The net catalytic turnover of BluB appears to be very different from IYD, and BluB has been recently identified as the source of the lower ligand (5,6-dimethylbenzimidazole) of vitamin B_{12} by sacrificing its bound FMN cofactor to oxidation (*116*). Still IYD and BluB form similar dimer interfaces at their core with criss-crossing helices and similar domain

swaps with their extended N- and C-terminal fingers (Supplementary Fig. S2). Upon binding substrate, MIT/DIT for IYD and molecular oxygen for BluB, the lid forms holding in the substrate near FMN. The differences between the IYD and BluB active site lid positioning may determine the specificity and explain the difference in function (Fig. 9.2). Blub active site lid caves in much closer to the FMN molecule, which would not be enough room for MIT or DIT to fit in the active site.

9.5B: Interface Structure of the Homodimer

For IYD, N- and C-terminal extensions of each polypeptide wrap around the other. These extensions span distances of greater than 37 Å and comprise a minimum of 26 amino acids near the N-terminus and 18 amino acids at the C-terminus. Two equivalent active sites are located within the dimer interface, and each active site is comprised of residues from both subunits. Accordingly subunit association is essential for FMN binding and catalytic activity. Because the IYD structure was solved as a monomer per asymmetric unit, its dimer was generated by crystallographic symmetry, and consequently no differences between the two monomers were observed. Differences observed between the monomeric units of IYD·MIT and IYD·DIT were minimal as evident from their root RMSD values of 0.196 Å and 0.254 Å, respectively.

9.5C: Substrate-induced Conformational Changes in the Active Site

Two unstructured regions were identified within substrate-free IYD by the lack of electron density corresponding to residues 156–177 and 195–208. Even in the substrate-bound co-crystals, the regions containing residues 195–208 exhibit

relatively weak electron density compared with other regions of the molecule. Only portions of this region, some with higher than average B-factors, could be built for IYD bound to substrate.

In contrast, residues 156–177 gain detectable structure based on the additional electron density observed for the co-crystals IYD·MIT and IYD·DIT. Substrate binding appears to induce the extension of one helix (αC) and the formation of another short helix (αD) in this region to cover the active site and protect substrate and flavin from solvent (Fig. 9.1B, D). This active site lid is stabilized in part by numerous interactions with the bound substrate that is in turn anchored by aromatic stacking with the isoalloxazine ring of FMN as well as by polar contacts with FMN and Ala-126 (Fig. 9.1E). Eight residues of the lid assemble within 4 Å of the substrate, and three of these form polar contacts (Glu-153, Tyr-157, and Lys-178).

An analysis of sequence conservation onto the surface of IYD indicates a high level of conservation on one face of the molecule, but not the other. I have designated those as front and back faces for analysis. Figure 9.3A shows the conserved face (front) of the protein along with solvent accessible FMN. As expected, there is significant residue conservation near the active site, and this conservation extends across this face. It is assumed that IYD requires interaction with a currently unknown reductase that binds and reduces the FMN. Therefore, the remaining conservation on the front face may be required for interaction with the reductase. The reductase would presumably bind to IYD and FMN when ligand is not present. Figure 9.3B displays the conserved surface when MIT is bound and the helical flap encloses the active site. The disappearance of the majority of the conserved surface supports the idea that the

conserved surface would only be required and open for binding when the substrate is absent.

Electrostatic comparisons between IYD·Holo and ligand bound IYD·MIT show distinct structural changes, both in shape and electrostatic charge density on the front face of the protein in response to substrate binding (Fig. 9.1C, D). It is clear that upon binding ligand the protein becomes more structured around the active site and the face of the protein becomes more defined with hydrophobic residues near the center and postively charged residues around the edges. This analysis also highlights that there is ample solvent accessibility to the FMN in the IYD·Holo to allow for substrate binding and interaction with a reductase.

Both MIT and DIT are stabilized in the active site by electrostatic interactions, expected by the conserved positively charged arginine residues near the phosphate tail of FMN and the negatively charged area surrounding the amino group of the substrate (Fig. 9.4A). The active site lid forms locking in the substrate near the FMN cofactor and provides just enough space for the specific substrate to stack on top of the FMN molecule. There is a slight conformational change of three nearby residues (Leu-169, Thr-174 and Leu-172) to allow enough room for DIT to fit into the active site (Fig. 9.4B).

9.6: Structural Basis for Deficiency of IYD in Humans

Moreno et al. discovered four mutations within the IYD gene from patients suffering from hypothyroidism (*106*). All mutations discovered are conserved in the mouse IYD gene, along with over 90% sequence identity between mouse and human homologs. This sequence conservation between human IYD and mouse IYD allows

for the acceptable explanation of the phenotypes expressed by the human mutations using the mouse IYD crystal structures. Mutations discovered at Arg101 to Trp (mouse: Arg97) and Phe105 and Ile106 to Leu (mouse: Phe101-Ile102) were observed in patients, and shown to completely abolish deiodinase activity (*106*). Arg97 is located at the rear of the active site, forming hydrogen bond to the phosphate group on the FMN cofactor (Fig. 9.5). Mutation of Arg97 to tryptophan would leave no room for FMN to bind correctly, resulting in the abolished activity. Phe101-Ile102 are also located near the active site (Fig. 9.5). Neighboring residue Arg100 extends and interacts with FMN intimately, with interactions to the 2'-ribityl hydroxyl group, N-1 and O-2. Mutations of Phe101-Ile102 to leucines would change the local structure of the loop and eliminate Arg100 interaction with FMN (Fig. 9.5). The fourth mutation, Ile116 (mouse: Ile112) to Thr was found in patients but still exhibited some deiodinase activity. Ile116 is not located near the active site but is buried within the protein in a hydrophobic core, clear from any interactions of the active site (Fig 9.5). Introducing threonine into the hydrophobic environment could create instability in the protein and could explain the decreased activity, albeit less severe.

9.7: IYD Acknowledgements

This work is based upon research conducted in part at the Northeastern Collaborative Access Team beamlines of the Advanced Photon Source supported by Award RR-15301 from the National Center for Research Resources at the National Institutes of Health. Use of the Advanced Photon Source is supported by the United States Department of Energy, Office of Basic Energy Sciences, under Contract DE-

AC02-06CH11357. We are grateful to the excellent staff at the Northeastern Collaborative Access Team for assistance with data collection and analysis.

Table 9.1: ShelX Results from HKL2MAP Program Suite.

The nine sulfur atoms are identified as shown above the line in the ShelXE results logfile, with the strength of anomalous signal present shown in the ShelXC results logfile (*111*, *112*).

ShelXC Results : Anomalous Signal

Resolution (Å)	>8.0	6.0	5.0	4.0	3.5	3.0	2.8	2.6	2.4	2.2	2.0
N(data)	337	436	546	1180	1202	2098	1324	1719	2354	3266	4674
<I/sig>	131.5	130.2	128.3	130.6	121.0	101.4	77.3	65.0	47.7	32.1	11.8
%Complete	96.8	99.5	99.8	99.8	99.8	99.9	99.9	99.9	99.9	100.0	99.4
<d"/sig>	2.05	2.07	2.08	1.47	1.42	1.33	1.23	1.12	1.10	1.04	1.01

ShelXE Results: Heavy Atoms

Site	x	y	z	Occupancy	Density
1	0.1561	0.6873	0.0164	87.7600	29.72
2	-0.0324	0.4312	-0.0509	76.3688	26.17
3	0.2386	0.5941	0.1696	66.2676	24.55
4	0.1652	0.3482	-0.0439	57.9567	18.45
5	0.3577	0.5715	0.1343	55.3502	19.80
6	0.3179	0.5090	0.0005	52.0943	15.01
7	0.0865	0.4359	0.1337	49.9618	18.33
8	0.1186	0.7462	0.1546	34.8583	10.25
9	0.0591	0.4298	0.0971	29.7945	10.94
10	0.3099	0.4724	-0.1720	19.5090	4.50
11	0.2347	0.5894	0.2019	16.6568	3.94
12	-0.0236	0.1274	0.0961	16.1654	2.58
13	0.2141	0.5821	0.0356	14.0679	2.22
14	-0.1755	-0.0113	0.2012	11.2421	3.80
15	0.4047	0.4973	0.1833	2.2554	2.39

Table 9.2: Data Collection and Refinement Statistics

Statistics for last shell are shown in parenthesis.

	IYD	**IYD·MIT**	**IYD·DIT**
PDB code	3GB5	3GFD	3GH8
Data collection			
Space group	$P3_12_1$	$P3_22_1$	$P2_1$
Cell dimensions			
a, *b*, *c* (Å)	87.76, 87.76, 62.65	105.19, 105.19, 162.12	50.61, 112.57, 189.25
α, β, γ (°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 89.92, 90.00
Molecules/Asymmetric unit	1	2	8
Wavelength (Å)	1.653	0.9795	0.9795
Resolution (Å)	50–2.0	50–2.45	50-2.6
R_{sym} (last shell)	0.057 (0.581)	0.081 (0.439)	0.086 (0.295)
$I/\sigma I$	63.7 (4.3)	20.0 (3.7)	9.7 (2.4)
Completeness (%)	99.8 (99.0)	99.2 (99.3)	97.1 (92.9)
Redundancy	14.0 (8.9)	4.7 (4.7)	2.6 (2.5)
Refinement			
Resolution (Å)	30-2.0	30–2.45	30–2.6
R_{work}/R_{free} (%)	16.7/18.9	14.3/17.8	18.1/26.4
No. protein residues monomers A/B/C/D/E/F/G/H	185	220/219	221/220/221/220/221 /220/221/220
No. non protein atoms			
Ligand	43	102	370
Solvent	132	540	300
Mean B-factors $(Å)^2$ (TLS refined)	32.6	38.8	35.5
RMS deviations			
Bond lengths (Å)	0.020	0.019	0.019
Bond angles (°)	1.569	2.016	1.926
Ramachandran plot			
Most favorable	92.5%	92.2%	88.3%
Additionally allowed	7.5%	7.3%	10.9%
Generously allowed	0.0%	0.5%	0.3%
Disallowed	0.0%	0.0%	0.4%

Table 9.3: IYD-Holo Scalepack Logfile

Shell		Average Redundancy Per Shell
Lower limit	Upper limit	
50.00	4.31	14.9
4.31	3.42	14.5
3.42	2.99	14.9
2.99	2.71	14.9
2.71	2.52	14.8
2.52	2.37	14.7
2.37	2.25	14.6
2.25	2.15	14.5
2.15	2.07	13.4
2.07	2.00	8.9
All	hkl	14.0

Shell		I/Sigma in resolution shells:								
Lower	Upper	% of reflections with I/Sigma less than								
limit	limit	0	1	2	3	5	10	20	>20	total
50.00	4.31	0.1	0.2	0.3	0.4	0.5	1.0	1.6	98.3	99.8
4.31	3.42	0.1	0.2	0.4	0.6	1.1	2.0	4.0	96.0	100.0
3.42	2.99	0.3	0.8	1.1	1.5	2.2	4.0	7.5	92.5	100.0
2.99	2.71	0.6	1.0	1.8	2.9	4.2	7.5	15.4	84.6	100.0
2.71	2.52	0.5	1.3	2.6	4.0	6.6	14.4	25.9	74.1	100.0
2.52	2.37	1.2	2.6	4.9	7.1	12.3	22.2	38.7	61.3	100.0
2.37	2.25	1.9	3.7	6.5	9.6	14.8	28.0	50.4	49.6	100.0
2.25	2.15	2.7	5.8	10.9	16.0	25.0	43.3	70.4	29.6	100.0
2.15	2.07	4.6	10.7	19.6	28.6	42.5	65.5	88.3	11.7	100.0
2.07	2.00	8.4	20.7	36.6	49.7	67.6	88.4	97.7	1.3	99.0
All	hkl	2.0	4.7	8.5	12.0	17.7	27.6	39.9	59.9	99.9

Shell			Average		Norm.	Linear	Square
Lower Limit	Upper Limit	I	error	stat.	Chi**2	R-fac	R-fac
50.00	4.31	795.1	7.7	3.9	1.665	0.030	0.035
4.31	3.42	588.6	10.1	3.3	1.461	0.052	0.060
3.42	2.99	275.3	3.8	2.2	1.985	0.056	0.062
2.99	2.71	128.8	2.0	1.6	1.861	0.065	0.069
2.71	2.52	87.9	1.8	1.5	1.700	0.082	0.083
2.52	2.37	55.8	1.6	1.5	1.600	0.115	0.112
2.37	2.25	43.6	1.6	1.6	1.610	0.153	0.148
2.25	2.15	28.1	1.6	1.6	1.525	0.232	0.218
2.15	2.07	17.0	1.8	1.8	1.434	0.374	0.348
2.07	2.00	10.7	2.5	2.5	1.208	0.581	0.540
All	reflections	203.8	3.5	2.2	1.626	0.057	0.049

Table 9.4: IYD·MIT Scalepack Logfile

Shell Lower limit	Upper limit	Average Redundancy Per Shell
50.00	5.28	4.5
5.28	4.19	4.6
4.19	3.66	4.7
3.66	3.32	4.7
3.32	3.09	4.7
3.09	2.90	4.7
2.90	2.76	4.7
2.76	2.64	4.7
2.64	2.54	4.7
2.54	2.45	4.7
All	hkl	4.7

Shell		I/Sigma in resolution shells:								
Lower	Upper	% of reflections with I/Sigma less than								
limit	limit	0	1	2	3	5	10	20	>20	total
50.00	5.28	0.6	1.5	2.3	3.3	5.3	9.8	22.2	76.3	98.4
5.28	4.19	1.2	2.3	3.5	4.8	6.7	12.7	28.4	70.3	98.7
4.19	3.66	1.4	3.2	5.3	7.1	10.4	19.7	38.7	60.6	99.3
3.66	3.32	1.3	3.7	7.5	10.7	16.6	30.3	56.2	43.2	99.4
3.32	3.09	3.4	7.3	12.1	16.9	25.2	42.2	67.7	31.6	99.3
3.09	2.90	6.2	12.4	19.6	26.5	39.0	59.2	82.5	16.9	99.5
2.90	2.76	7.1	16.3	26.2	35.9	49.0	70.8	90.0	9.3	99.3
2.76	2.64	8.6	19.7	31.8	41.6	57.4	78.4	92.5	6.6	99.1
2.64	2.54	11.5	24.4	39.0	50.1	66.1	84.4	96.3	2.9	99.2
2.54	2.45	13.1	28.6	45.1	57.6	72.3	88.6	96.6	2.7	99.3
All	hkl	5.4	11.8	19.0	25.2	34.4	49.1	66.6	32.5	99.2

Shell			Average		Norm.	Linear	Square
Lower Limit	Upper Limit	I	error	stat.	Chi**2	R-fac	R-fac
50.00	5.28	103.9	2.9	1.8	1.182	0.034	0.034
5.28	4.19	105.8	3.5	2.2	1.246	0.051	0.058
4.19	3.66	81.7	2.8	2.1	1.678	0.066	0.068
3.66	3.32	50.1	2.4	2.0	1.138	0.078	0.076
3.32	3.09	33.5	1.9	1.9	1.010	0.093	0.082
3.09	2.90	19.0	1.8	1.7	0.974	0.150	0.126
2.90	2.76	13.1	1.7	1.7	0.942	0.213	0.182
2.76	2.64	10.8	1.8	1.7	0.898	0.258	0.209
2.64	2.54	8.0	1.8	1.7	0.884	0.357	0.300
2.54	2.45	6.7	1.8	1.8	0.905	0.439	0.346
All	reflections	43.9	2.2	1.9	1.086	0.081	0.062

Table 9.5: IYD·DIT Scalepack Logfile

Shell Lower limit	Upper limit	Average Redundancy Per Shell
50.00	5.60	2.6
5.60	4.45	2.6
4.45	3.88	2.6
3.88	3.53	2.6
3.53	3.28	2.6
3.28	3.08	2.6
3.08	2.93	2.6
2.93	2.80	2.6
2.80	2.69	2.6
2.69	2.60	2.5
All	hkl	2.6

Shell		I/Sigma in resolution shells:								
Lower	Upper	% of reflections with I/Sigma less than								
limit	limit	0	1	2	3	5	10	20	>20	total
50.00	5.60	0.9	2.5	4.5	6.5	10.0	22.4	77.0	18.1	95.1
5.60	4.45	1.0	3.2	5.8	8.5	14.0	30.5	83.8	13.0	96.8
4.45	3.88	1.3	3.8	7.2	10.5	17.1	36.2	86.4	11.0	97.4
3.88	3.53	2.2	6.9	12.6	18.3	28.1	50.4	91.5	6.3	97.9
3.53	3.28	3.3	10.2	18.6	26.4	39.1	61.3	90.9	7.3	98.2
3.28	3.08	5.9	15.4	27.0	36.3	49.9	71.6	94.4	4.1	98.5
3.08	2.93	9.5	22.5	37.5	48.1	61.9	80.9	96.1	2.7	98.9
2.93	2.80	12.4	29.1	45.4	56.6	69.8	86.1	97.7	1.1	98.8
2.80	2.69	13.5	31.6	48.1	58.9	71.4	86.1	95.7	1.2	96.9
2.69	2.60	14.1	33.3	51.1	62.1	73.2	85.0	91.4	1.5	92.9
All	hkl	6.4	15.8	25.7	33.1	43.3	60.9	90.5	6.7	97.1

Shell			Average		Norm.	Linear	Square
Lower Limit	Upper Limit	I	error	stat.	Chi**2	R-fac	R-fac
50.00	5.60	154.2	9.2	4.4	0.817	0.039	0.041
5.60	4.45	128.0	8.5	4.9	1.109	0.060	0.065
4.45	3.88	135.1	9.4	5.8	1.122	0.063	0.068
3.88	3.53	88.7	7.7	5.7	1.036	0.076	0.072
3.53	3.28	61.2	6.7	5.8	1.076	0.096	0.088
3.28	3.08	42.4	6.3	5.8	1.049	0.124	0.109
3.08	2.93	26.5	5.9	5.7	1.053	0.178	0.154
2.93	2.80	19.1	5.9	5.7	1.004	0.236	0.209
2.80	2.69	17.1	5.9	5.8	1.058	0.258	0.223
2.69	2.60	14.1	5.9	5.9	0.973	0.295	0.239
All	reflections	68.8	7.1	5.6	1.030	0.086	0.070

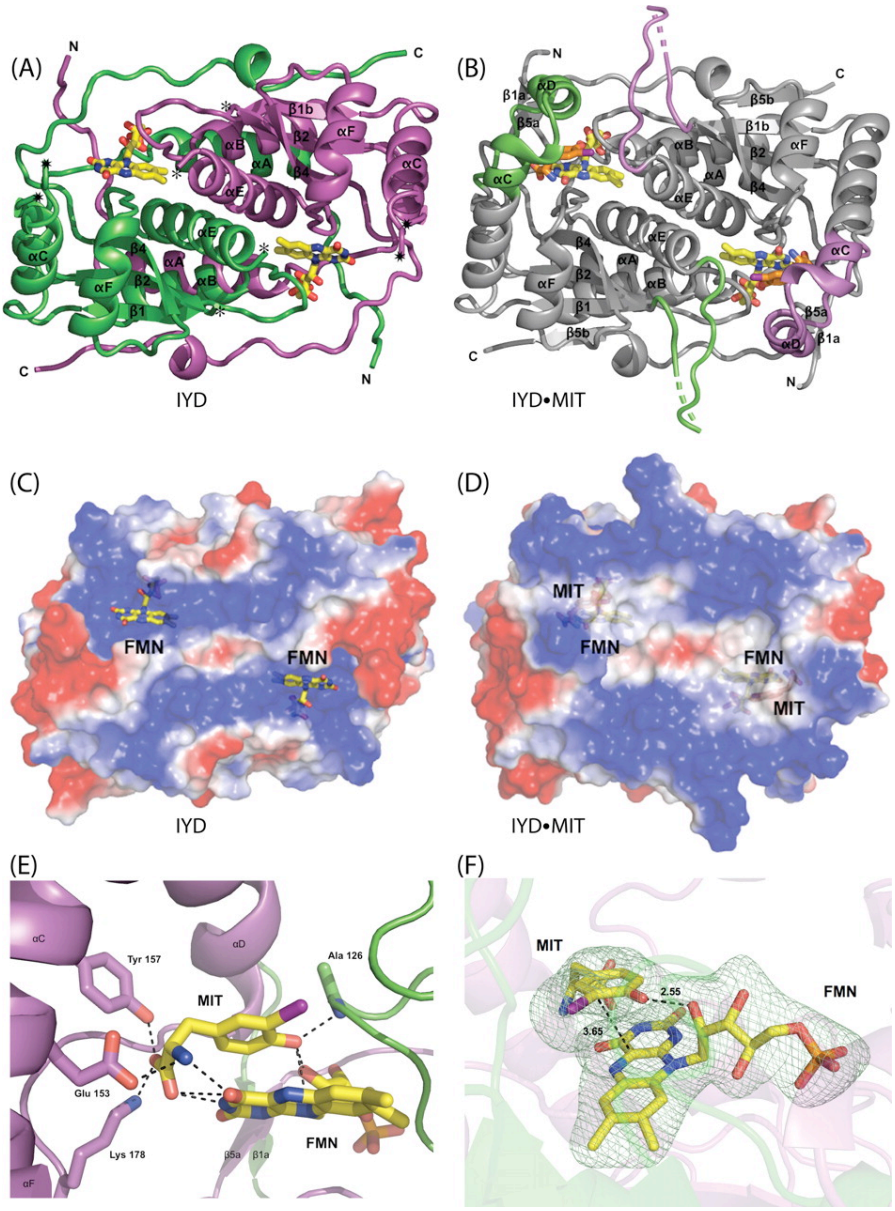


Figure 9.1: IYD structure

A) An overall view of the native homodimer of IYD crystallized in the absence of substrate. Each monomer is distinguished by color. Disordered regions consisting of residues 156–177 and 195–208 connect to the structure as indicated by ∗ and ✱, respectively. **B)** Native homodimer of IYD crystallized in the presence of its substrate, MIT. Only the structure induced upon substrate binding is highlighted in

the colors of the monomers shown in (**A**). The surface properties of IYD (**C**) and its complex with monoiodotyrosine (**D**) were calculated using vacuum electrostatics in PyMOL (*75*). Blue indicates positive charge, and red indicates negative charge. **E)** Ionic interactions and hydrogen bonding stabilize the FMN·monoiodotyrosine complex formed by IYD. **F)** The interaction between FMN and MIT in the active site of IYD. An $F_o - F_c$ electron density map calculated after refinement in the absence of FMN and MIT is shown contoured at 3σ.

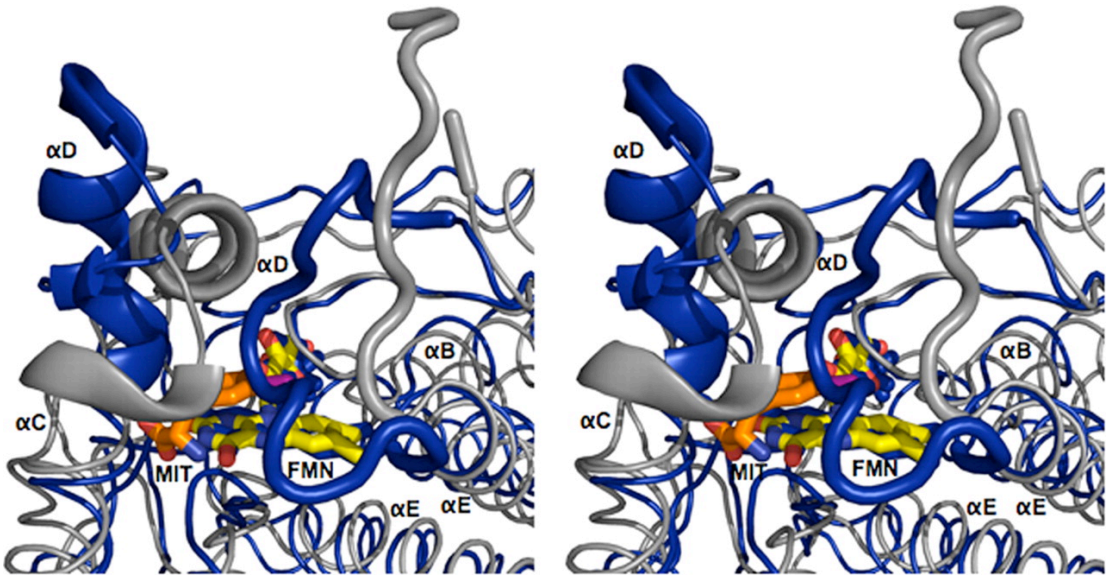


Figure 9.2: Structural Overlay of IYD and BluB

Structural differences between IYD·MIT (gray) and BluB (blue) are highlighted using an overlaid stereoimage. Structural changes near the active site are depicted with the schematic representation. The full-view overlay can be seen as a stereoimage in the supplemental information.

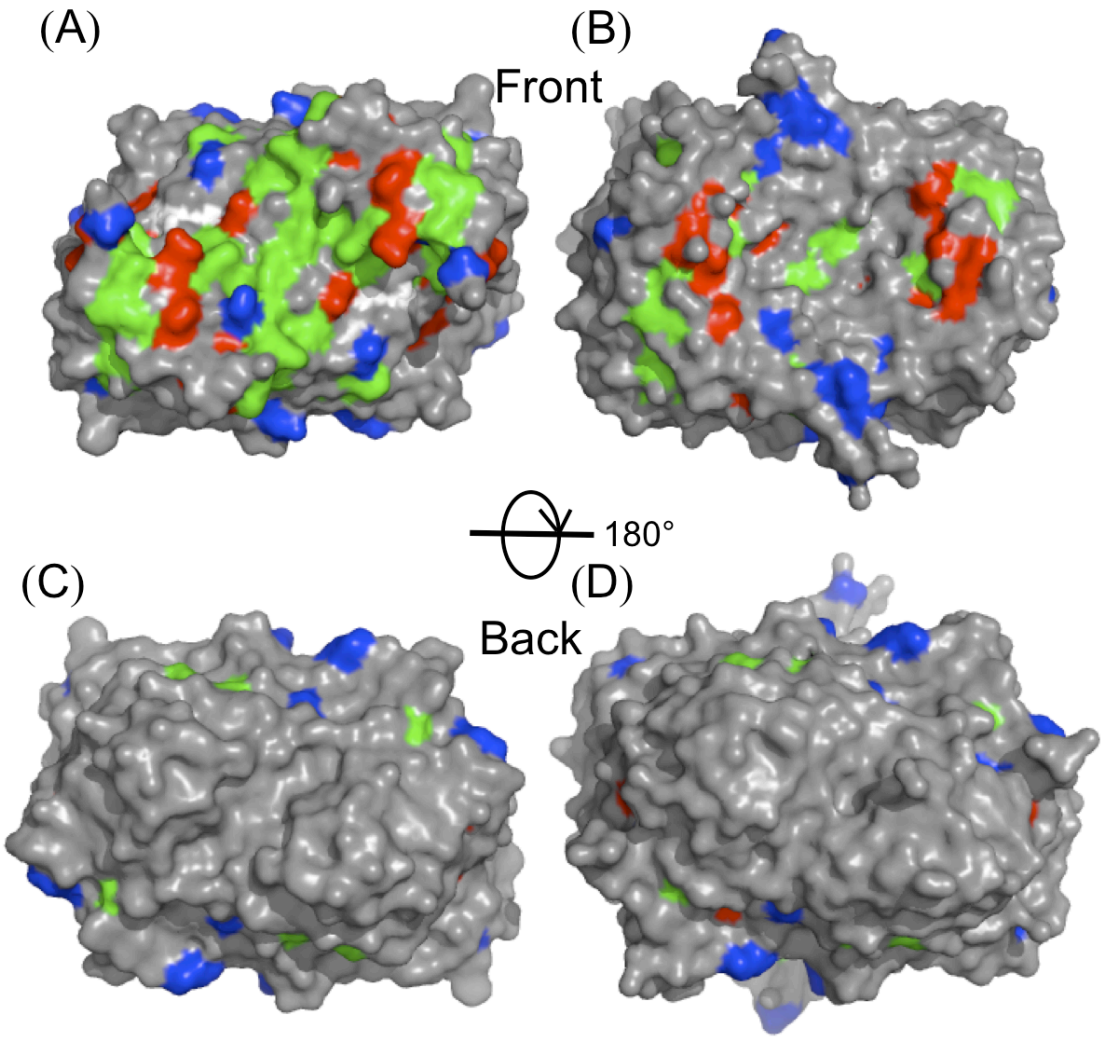


Figure 9.3: IYD Sequence Conservation

A) and **B)** shows front views of sequence conservation modeled onto the surface of IYD·Holo and IYD·MIT, respectively. Red=sequence identity. Green=strongly similar. Blue=weakly similar. Gray=no similarity White=FMN surface. **C)** and **D)** are rotated 180˚ around the x-axis in respect to (**A**) and (**B**). ClustalW used for alignment from the following: M. musculus (mouse), X. laevis (frog), D. rerio (zebra fish), N. vectinsus (sea anemone), S. purpuratus (sea urchin), D. melengaster (fruit fly), and C. elegans (worm).

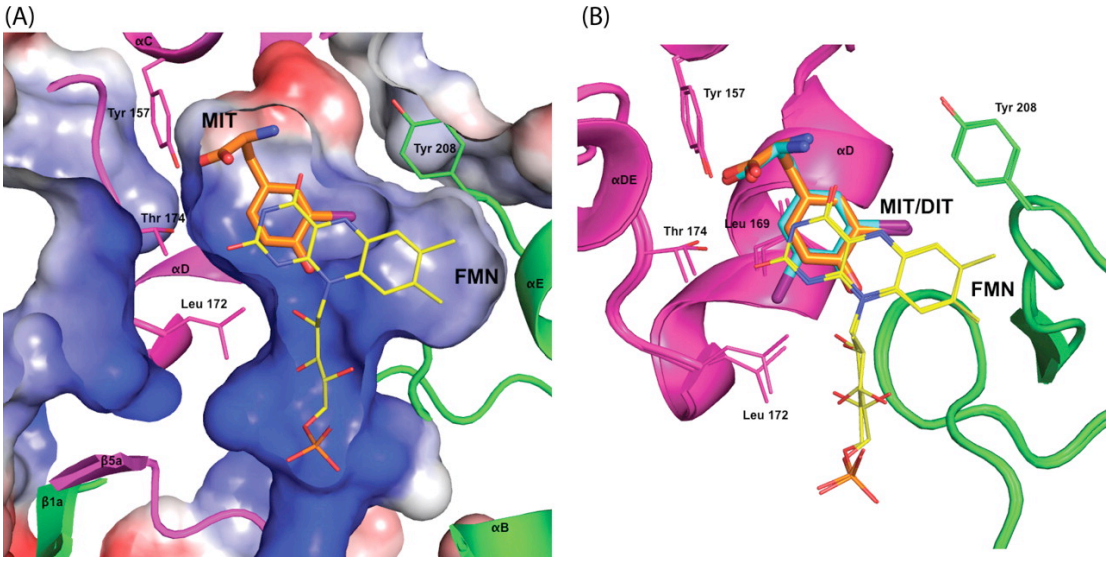


Figure 9.4: IYD Active Site Electrostatics and MIT/DIT Alignment

A) The surface characteristics of the active site of IYD for the IYD·MIT co-crystal crystal calculated using vacuum electrostatics in PyMOL (47). Blue indicates positive charge, and red indicates negative charge. **B)** Alignment of active site structures of IYD bound with MIT (orange) and DIT (cyan) illustrates the minor conformational change required to accommodate the larger substrate.

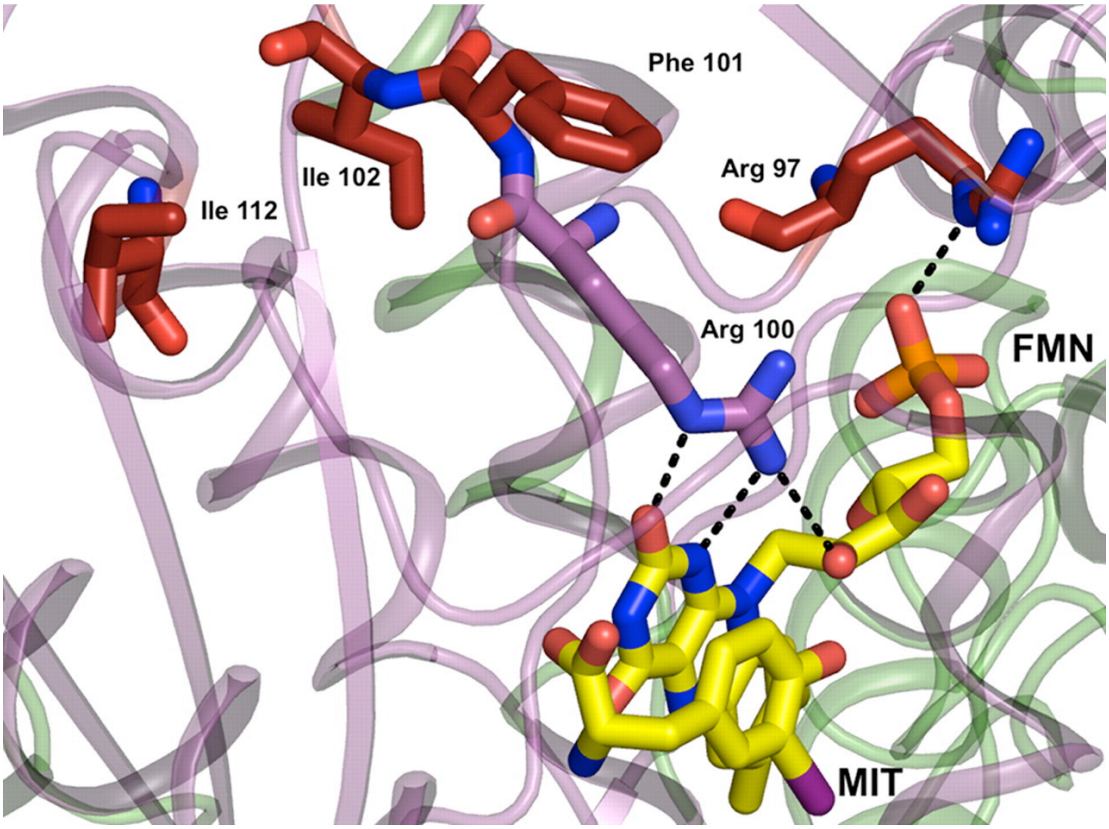


Figure 9.5: Mapping Human Mutations onto the Structure of IYD

Native residues of IYD (M. musculus) highlighted in red correspond to sites associated with human mutations identified clinically to cause hypothyroidism (*106*). Other color coding is consistent with the previous illustrations (see legends) and distinguishes the two subunits within the dimer and the FMN.

SUPPLEMENTAL FIGURES FOR

CRYSTAL STRUCTURE OF IODOTYROSINE DEIODINASE, A NOVEL FLAVOPROTEIN RESPONSIBLE FOR IODIDE SALVAGE IN THYROID GLANDS

Seth R. Thomas[1,2], Patrick M. McTamney[1], Jennifer M. Adler[1],
Nicole LaRonde-LeBlanc[1,2] and Steven E. Rokita[1],
From the Department of Chemistry and Biochemistry[1] and the Center for Biomolecular Structure and Organization[2],
University of Maryland, College Park, Maryland 20742 USA

Address correspondence to Nicole LaRonde-LeBlanc, Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742 USA. Fax 301-314-0386; E-mail: nlaronde@umd.edu and Steven E. Rokita Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742 USA. Fax 301-405-9376; E-mail: rokita@umd.edu

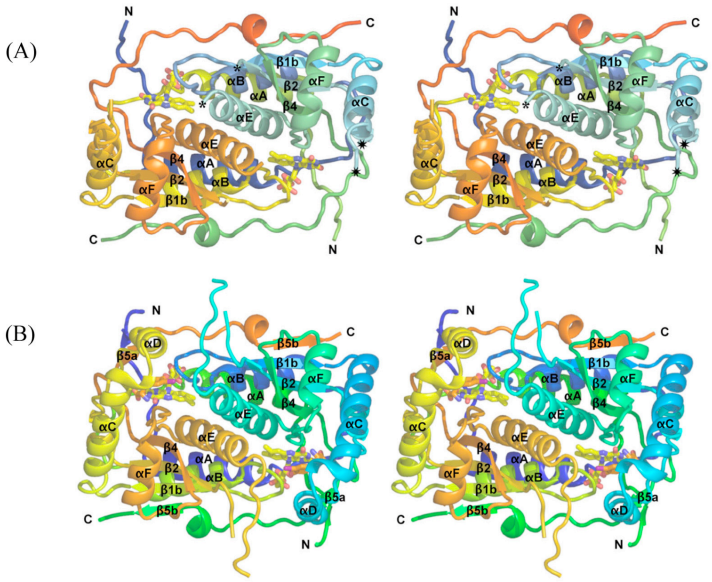


Figure S1. (A) An overall stereo view of IYD and (B) IYD·MIT colored to differentiate between secondary structural elements.

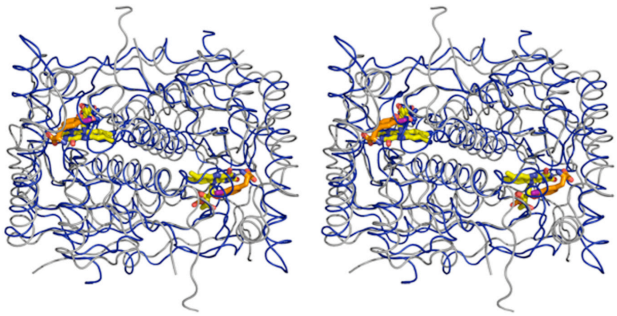


Figure S2 A stereo view of the structural alignment of IYD·MIT (grey) and BluB (blue).

Bibliography

1. Fatica, A., and Tollervey, D. (2002) Making ribosomes. *Current Opinion in Cell Biology 14*, 313-318.

2. Venema, J., and Tollervey, D. (1999) Ribosome synthesis in saccharomyces cerevisiae. *Annual Review of Genetics 33*, 261-311.

3. Fromont-Racine, M., Senger, B., Saveanu, C., and Fasiolo, F. (2003) Ribosome assembly in eukaryotes. *Gene 313*, 17-42.

4. Kiss, T. (2001) Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO 20*, 3617-3622.

5. Maden, B. E. H. (1990) The numerous modified nucleotides in eukaryotic ribosomal RNA. In *Progress in Nucleic Acid Research and Molecular Biology* (Waldo, E. C., and Kivie, M., Eds.), pp 241-303, Academic Press.

6. Rozenski, J., Crain, P. F., and McCloskey, J. A. (1999) The RNA modification database: 1999 update. *Nucleic Acids Research 27*, 196-197.

7. Decatur, W. A., and Fournier, M. J. (2002) rRNA modifications and ribosome function. *Trends in Biochemical Sciences 27*, 344-351.

8. Kiss-Laszlo, Z., Henry, Y., Bachellerie, J. P., Caizergues-Ferrer, M., and Kiss, T. (1996) Site-specific ribose methylation of preribosomal RNA: A novel function for small nucleolar RNAs. *Cell 85*, 1077 - 1088.

9. Kressler, D., Linder, P., and de la Cruz, J. (1999) Protein trans-acting factors involved in ribosome biogenesis in Saccharomyces cerevisiae. *Molecular and Cellular Biology 19*, 7897-7912.

10. Kendrew, J., Dodo, G., Dintzis, H., Parrish, R., Wyckoff, H., and Phillips, D. (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature 181*, 662-666.

11. Williamson, M. P., Havel, T. F., and W¸Thrich, K. (1985) Solution conformation of proteinase inhibitor iia from bull seminal plasma by 1H

nuclear magnetic resonance and distance geometry. *Journal of Molecular Biology 182*, 295-315.

12. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The protein data bank. *Nucleic Acids Research 28*, 235-242.

13. Hakuno, F., Hughes, D. A., and Yamamoto, M. (1996) The Schizosaccharomyces pombe Mra1 gene, which is required for cell growth and mating, can suppress the mating inefficiency caused by a deficit in the Ras1 activity. *Genes to Cells 1*, 303-315.

14. Liu, P. C. C., and Thiele, D. J. (2001) Novel stress-responsive genes Emg1 and Nop14 encode conserved, interacting proteins required for 40S ribosome biogenesis. *Molecular Biology of the Cell 12*, 3644-3657.

15. Eschrich, D., Buchhaupt, M., Kotter, P., and Entian, K.-D. (2002) Nep1p (Emg1p), a novel protein conserved in eukaryotes and archaea, is involved in ribosome biogenesis. *Current Genetics 40*, 326-338.

16. Taylor, A. B., Meyer, B., Leal, B. Z., Kotter, P., Schirf, V., Demeler, B., Hart, P. J., Entian, K. D., and Wohnert, J. (2008) The crystal structure of Nep1 reveals an extended SPOUT-class methyltransferase fold and a pre-organized SAM-binding site. *Nucleic Acids Research 36*, 1542-1554.

17. Leulliot, N., Bohnsack, M. T., Graille, M., Tollervey, D., and Van Tilbeurgh, H. (2007) The yeast ribosome synthesis factor Emg1 is a novel member of the superfamily of alpha/beta knot fold methyltransferases. *Nucleic Acids Research 36*, 629-639.

18. Anantharaman, V., Koonin, E. V., and Aravind, L. (2002) SPOUT: A class of methyltransferases that includes Spou and Trmd RNA methylase superfamilies, and novel superfamilies of predicted prokaryotic RNA methylases. *Journal of Molecular Microbiology and Biotechnology 4*, 71 - 75.

19. Tkaczuk, K. L., Dunin-Horkawicz, S., Purta, E., and Bujnicki, J. M. (2007) Structural and evolutionary bioinformatics of the SPOUT superfamily of methyltransferases. *BMC Bioinformatics 8*, 73.

20. Ero, R., Peil, L., Liiv, A., and Remme, J. (2008) Identification of pseudouridine methyltransferase in Escherichia coli. *RNA 14*, 2223-2233.

21. Buchhaupt, M., Meyer, B., Kötter, P., and Entian, K.-D. (2006) Genetic evidence for 18S rRNA binding and an Rps19p assembly function of yeast nucleolar protein Nep1p. *Molecular Genetics and Genomics 276*, 273-284.

22. Wurm, J. P., Meyer, B., Bahr, U., Held, M., Frolow, O., Kotter, P., Engels, J. W., Heckel, A., Karas, M., Entian, K. D., and Wohnert, J. (2010) The ribosome assembly factor Nep1 responsible for Bowen-Conradi syndrome is a pseudouridine-N1-specific methyltransferase. *Nucleic Acids Research 38*, 2387-2398.

23. Bakin, A., and Ofengand, J. (1995) Mapping of the 13 pseudouridine residues in Saccharomyces cerevisiae small subunit ribosomal RNA to nucleotide resolution. *Nucleic Acids Research 23*, 3290-3294.

24. Draptchinskaia, N., Gustavsson, P., Andersson, B., Pettersson, M., Willig, T.-N., Dianzani, I., Ball, S., Tchernia, G., Klar, J., Matsson, H., Tentler, D., Mohandas, N., Carlsson, B., and Dahl, N. (1999) The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. *Nature Genetics 21*, 169-175.

25. Morimoto, K., Lin, S., and Sakamoto, K. (2007) The functions of Rps19 and their relationship to Diamond–Blackfan anemia: A review. *Molecular Genetics and Metabolism 90*, 358-362.

26. Ferreiracerca, S., Poll, G., Gleizes, P., Tschochner, H., and Milkereit, P. (2005) Roles of eukaryotic ribosomal proteins in maturation and transport of pre-18S rRNA and ribosome function. *Molecular Cell 20*, 263-275.

27. Idol, R., Robledo, S., Du, H., Crimmins, D., Wilson, D., Ladenson, J., Bessler, M., and Mason, P. (2007) Cells depleted for Rps19, a protein associated with Diamond Blackfan anemia, show defects in 18S ribosomal RNA synthesis and small ribosomal subunit production. *Blood Cells, Molecules, and Diseases 39*, 35-43.

28. Leger-Silvestre, I. (2005) Specific role for yeast homologs of the Diamond Blackfan anemia-associated Rps19 protein in ribosome synthesis. *Journal of Biological Chemistry 280*, 38177-38185.

29. Lowe, T. M. (1999) A computational screen for methylation guide snoRNAs in yeast. *Science 283*, 1168-1171.

30. Taylor, D. J., Devkota, B., Huang, A. D., Topf, M., Narayanan, E., Sali, A., Harvey, S. C., and Frank, J. (2009) Comprehensive molecular structure of the eukaryotic ribosome. *Structure 17*, 1591-1604.

31. Hunter, T. (2000) Signaling--2000 and beyond. *Cell 100*, 113-127.

32. Ptacek, J., Devgan, G., Michaud, G., Zhu, H., Zhu, X., Fasolo, J., Guo, H., Jona, G., Breitkreutz, A., Sopko, R., McCartney, R. R., Schmidt, M. C., Rachidi, N., Lee, S.-J., Mah, A. S., Meng, L., Stark, M. J. R., Stern, D. F., De Virgilio, C., Tyers, M., Andrews, B., Gerstein, M., Schweitzer, B., Predki, P. F., and Snyder, M. (2005) Global analysis of protein phosphorylation in yeast. *Nature 438*, 679-684.

33. Angermayr, M., Roidl, A., and Bandlow, W. (2002) Yeast Rio1p is the founding member of a novel subfamily of protein serine kinases involved in the control of cell cycle progression. *Molecular Microbiology 44*, 309-324.

34. Angermayr, M., and Bandlow, W. (2002) Rio1, an extraordinary novel protein kinase. *FEBS Letters 524*, 31-36.

35. LaRonde-LeBlanc, N., Guszczynski, T., Copeland, T., and Wlodawer, A. (2005) Structure and activity of the atypical serine kinase Fio1. *FEBS 272*, 3698-3713.

36. LaRonde-LeBlanc, N., Guszczynski, T., Copeland, T., and Wlodawer, A. (2005) Autophosphorylation of Archaeoglobus fulgidus Rio2 and crystal structures of its nucleotide-metal ion complexes. *FEBS 272*, 2800-2810.

37. LaRonde-LeBlanc, N., and Wlodawer, A. (2004) Crystal structure of A. Fulgidus Rio2 defines a new family of serine protein kinases. *Structure 12*, 1585-1594.

38. Geerlings, T. H., Faber, A. W., Bister, M. D., Vos, J. C., and Raue, H. A. (2003) Rio2p, an evolutionarily conserved, low abundant protein kinase essential for processing of 20S pre-rRNA in Saccharomyces cerevisiae. *Journal of Biological Chemistry 278*, 22537-22545.

39. LaRonde-LeBlanc, N., and Wlodawer, A. (2005) The Rio kinases: An atypical protein kinase family required for ribosome biogenesis and cell cycle progression. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics 1754*, 14-24.

40. Vanrobays, E., Gelugne, J.-P., Gleizes, P.-E., and Caizergues-Ferrer, M. (2003) Late cytoplasmic maturation of the small ribosomal subunit requires Rio proteins in Saccharomyces cerevisiae. *Molecular and Cellular Biology 23*, 2083-2095.

41. Vanrobays, E., Gleizes, P.-E., Bousquet-Antonelli, C., Noaillac-Depeyre, Caizergues-Ferrer, M., and Gelugne, J.-P. (2001) Processing of 20S pre-rRNA to 18S ribosomal RNA in yeast requires Rrp10p, an essential non-ribosomal cytoplasmic protein. *EMBO 20*, 4204-4213.

42. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002) The protein kinase complement of the human genome. *Science 298*, 1912-1934.

43. Hanks, S. K., and Hunter, T. (1995) Protein kinases 6. The eukaryotic protein kinase superfamily: Kinase (catalytic) domain structure and classification. *FASEB J. 9*, 576-596.

44. Hanks, S. K., Quinn, A. M., and Hunter, T. (1988) The protein kinase family: Conserved features and deduced phylogeny of the catalytic domains. *Science 241*, 42-52.

45. Knighton, D. R., Zheng, J. H., Ten Eyck, L. F., Xuong, N. H., Taylor, S. S., and Sowadski, J. M. (1991) Structure of a peptide inhibitor bound to the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science 253*, 414-420.

46. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Clustal w: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research 22*, 4673-4680.

47. Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V., Lescot, M., Claverie, J. M., and

Gascuel, O. (2008) Phylogeny.Fr: Robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research 36*, W465-W469.

48. LaRonde-LeBlanc, N., and Wlodawer, A. (2005) A family portrait of the rio kinases. *Journal of Biological Chemistry 280*, 37297-37300.

49. Kimmelman, A. C., Hezel, A. F., Aguirre, A. J., Zheng, H., Paik, J.-h., Ying, H., Chu, G. C., Zhang, J. X., Sahin, E., Yeo, G., Ponugoti, A., Nabioullin, R., Deroo, S., Yang, S., Wang, X., McGrath, J. P., Protopopova, M., Ivanova, E., Zhang, J., Feng, B., Tsao, M. S., Redston, M., Protopopov, A., Xiao, Y., Futreal, P. A., Hahn, W. C., Klimstra, D. S., Chin, L., and DePinho, R. A. (2008) Genomic alterations link Rho family of gtpases to the highly invasive phenotype of pancreas cancer. *Proceedings of the National Academy of Sciences 105*, 19372-19377.

50. Small, J. V., Stradal, T., Vignal, E., and Rottner, K. (2002) The lamellipodium: Where motility begins. *Trends in Cell Biology 12*, 112-120.

51. Cannone, J., Subramanian, S., Schnare, M., Collett, J., D'Souza, L., Du, Y., Feng, B., Lin, N., Madabusi, L., Muller, K., Pande, N., Shang, Z., Yu, N., and Gutell, R. (2002) The comparative RNA web (crw) site: An online database of comparative sequence and structure information for ribosomal, intron, and other rnas. *BMC Bioinformatics 3*, 2.

52. Stetter, K. O. (1988) Archaeoglobus fulgidus gen. Nov., sp. Nov.: A new taxon of extremely thermophilic archaebacteria. *Systematic and Applied Microbiology 10*, 172-173.

53. Stetter, K. O., Lauerer, G., Thomm, M., and Neuner, A. (1987) Isolation of extremely thermophilic sulfate reducers: Evidence for a novel branch of archaebacteria. *Science 236*, 822-824.

54. Beeder, J., Nilsen, R. K., Rosnes, J. T., Torsvik, T., and Lien, T. (1994) Archaeoglobus fulgidus isolated from hot north sea oil field waters. *Applied and Environmental Microbiology 60*, 1227-1231.

55. Kapust, R. B., TˆzsÈr, J., Copeland, T. D., and Waugh, D. S. (2002) The p1' specificity of tobacco etch virus protease. *Biochemical and Biophysical Research Communications 294*, 949-955.

56. Otwinowski, Z., Minor, W., and Charles W. Carter, Jr. (1997) Processing of x-ray diffraction data collected in oscillation mode. In *Methods in Enzymology*, pp 307-326, Academic Press.

57. Hendrickson, W. A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science 254*, 51-58.

58. Hendrickson, W. A., and Ogata, C. M. (1997) Phase determination from multiwavelength anomalous diffraction measurements. In *Methods in Enzymology* (Charles W. Carter, Jr., Ed.), pp 494-523, Academic Press.

59. Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G., and Sheldrick, G. M. (1999) Can anomalous signal of sulfur become a tool for solving protein crystal structures? *Journal of Molecular Biology 289*, 83-92.

60. Hendrickson, W. A., Horton, J. R., and LeMaster, D. M. (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): A vehicle for direct determination of three-dimensional structure. *Embo J 9*, 1665-1672.

61. Garman, E., and Murray, J. W. (2003) Heavy-atom derivatization. *Acta Crystallographica Section D 59*, 1903-1913.

62. Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for x-ray crystallography using Arp/Warp version 7. *Nature Protocols 3*, 1171-1179.

63. Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) Automated structure solution with autosharp. In *Macromolecular crystallography Protocols*, pp 215-230.

64. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D 60*, 2126-2132.

65. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D 53*, 240-255.

66. Collaborative. (1994) The ccp4 suite: Programs for protein crystallography. *Acta Crystallographica Section D 50*, 760-763.

67. Vagin, A., and Teplyakov, A. (1997) Molrep: An automated program for molecular replacement. *Journal of Applied Crystallography 30*, 1022-1025.

68. Vagin, A., and Teplyakov, A. (2000) An approach to multi-copy search in molecular replacement. *Acta Crystallographica Section D 56*, 1622-1624.

69. Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallographica Section D 59*, 1131-1137.

70. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *Journal of Applied Crystallography 40*, 658-674.

71. Adams, P. D., Afonine, P. V., Grosse-Kunstleve, R. W., Read, R. J., Richardson, J. S., Richardson, D. C., and Terwilliger, T. C. (2009) Recent developments in phasing and structure refinement for macromolecular crystallography. *Current Opinion in Structural Biology 19*, 566-572.

72. 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., and Zwart, P. H. (2010) Phenix: A comprehensive python-based system for macromolecular structure solution. *Acta Crystallographica Section D 66*, 213-221.

73. Painter, J., and Merritt, E. A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallographica Section D 62*, 439-450.

74. Beeby, M., O'Connor, B. D., Ryttersgaard, C., Boutz, D. R., Perry, L. J., and Yeates, T. O. (2005) The genomics of disulfide bonding and protein stabilization in thermophiles. *PLoS Biol 3*, e309.

75. Delano, W. (2002) The pymol molecular graphics system. *Delano Scientific*.

76. Ofengand, J., and Schaefer, H. (1965) On the ionization constant of 5-ribosyluracil. *Biochemistry 4*, 2832-2835.

77. Armistead, J., Khatkar, S., Meyer, B., Mark, B. L., Patel, N., Coghlan, G., Lamont, R. E., Liu, S., Wiechert, J., and Cattini, P. A. (2009) Mutation of a gene essential for ribosome biogenesis, Emg1, causes bowen-conradi syndrome. *The American Journal of Human Genetics 84*, 728-739.

78. Wurm, J., Duchardt, E., Meyer, B., Leal, B., Kötter, P., Entian, K.-D., and Wöhnert, J. (2009) Backbone resonance assignments of the 48 kDa dimeric putative 18S rRNA-methyltransferase Nep1 from Methanocaldococcus jannaschii. *Biomolecular NMR Assignments 3*, 251-254-254.

79. Hofacker, I. L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Research 31*, 3429-3431.

80. Alian, A., Lee, T. T., Griner, S. L., Stroud, R. M., and Finer-Moore, J. (2008) Structure of a trma:RNA complex: A consensus RNA fold contributes to substrate selectivity and catalysis in m5u methyltransferases. *Proceedings of the National Academy of Sciences 105*, 6876-6881.

81. Hur, S., Stroud, R. M., and Finer-Moore, J. (2006) Substrate recognition by RNA 5-methyluridine methyltransferases and pseudouridine synthases: A structural perspective. *Journal of Biological Chemistry 281*, 38969-38973.

82. Ben-Shem, A., Jenner, L., Yusupova, G., and Yusupov, M. (2010) Crystal structure of the eukaryotic ribosome. *Science 330*, 1203-1209.

83. Schafer, T., Maco, B., Petfalski, E., Tollervey, D., B√∂ttcher, B., Aebi, U., and Hurt, E. (2006) Hrr25-dependent phosphorylation state regulates organization of the pre-40S subunit. *Nature 441*, 651-655.

84. White, J., Li, Z., Sardana, R., Bujnicki, J. M., Marcotte, E. M., and Johnson, A. W. (2008) Bud23 methylates G1575 of 18S rRNA and is required for efficient nuclear export of pre-40S subunits. *Molecular and Cellular Biology 28*, 3151-3161.

85. Fenner, B. J., Scannell, M., and Prehn, J. H. M. (2009) Identification of polyubiquitin binding proteins involved in NF-κβ signaling using protein

arrays. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics 1794*, 1010-1016.

86. Zheng, J., Sagar, V., Smolinsky, A., Bourke, C., LaRonde-LeBlanc, N., and Cropp, T. A. (2009) Structure and function of the macrolide biosensor protein, Mphr(a), with and without erythromycin. *Journal of Molecular Biology 387*, 1250-1260.

87. Balcewich, M. D., Reeve, T. M., Orlikow, E. A., Donald, L. J., Vocadlo, D. J., and Mark, B. L. (2010) Crystal structure of the ampr effector binding domain provides insight into the molecular regulation of inducible Ampc β-lactamase. *Journal of Molecular Biology 400*, 998-1010.

88. Matthews, B. W. (1968) Solvent content of protein crystals. *Journal of Molecular Biology 33*, 491-497.

89. Kantardjieff, K. A., and Rupp, B. (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Science 12*, 1865-1871.

90. Harp, J. M., Hanson, B. L., Timm, D. E., and Bunick, G. J. (1999) Macromolecular crystal annealing: Evaluation of techniques and variables. *Acta Crystallographica Section D 55*, 1329-1334.

91. Heras, B., and Martin, J. L. (2005) Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallographica Section D 61*, 1173-1180.

92. Heras, B., Edeling, M. A., Byriel, K. A., Jones, A., Raina, S., and Martin, J. L. (2003) Dehydration converts Dsbg crystal diffraction from low to high resolution. *Structure (London, England : 1993) 11*, 139-145.

93. Bergfors, T. (2003) Seeds to crystals. *Journal of Structural Biology 142*, 66-76.

94. Stura, E. A., and Wilson, I. A. (1991) Applications of the streak seeding technique in protein crystallization. *Journal of Crystal Growth 110*, 270-282.

95. Cooper, D. R., Boczek, T., Grelewska, K., Pinkowska, M., Sikorska, M., Zawadzki, M., and Derewenda, Z. (2007) Protein crystallization by surface entropy reduction: Optimization of the SER strategy. *Acta Crystallographica Section D 63*, 636-645.

96. Longenecker, K. L., Garrard, S. M., Sheffield, P. J., and Derewenda, Z. S. (2001) Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. *Acta Crystallographica Section D 57*, 679-688.

97. Goldschmidt, L., Cooper, D. R., Derewenda, Z. S., and Eisenberg, D. (2007) Toward rational protein crystallization: A web server for the design of crystallizable protein variants. *Protein Science 16*, 1569-1576.

98. Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology 292*, 195 - 202.

99. Bryson, K., McGuffin, L. J., Marsden, R. L., Ward, J. J., Sodhi, J. S., and Jones, D. T. (2005) Protein structure prediction servers at University College London. *Nucleic Acids Research 33*, W36-W38.

100. Chen, Z., and Pickart, C. M. (1990) A 25-kilodalton ubiquitin carrier protein (E2) catalyzes multi-ubiquitin chain synthesis via lysine 48 of Ubiquitin. *Journal of Biological Chemistry 265*, 21835-21842.

101. Vijay-kumar, S., Bugg, C. E., and Cook, W. J. (1987) Structure of ubiquitin refined at 1.8 Å resolution. *Journal of Molecular Biology 194*, 531-544.

102. Ko, S., Kang, G. B., Song, S. M., Lee, J.-G., Shin, D. Y., Yun, J.-H., Sheng, Y., Cheong, C., Jeon, Y. H., Jung, Y.-K., Arrowsmith, C. H., Avvakumov, G. V., Dhe-Paganon, S., Yoo, Y. J., Eom, S. H., and Lee, W. (2010) Structural basis of E2-25K/UBB+1 interaction leading to proteasome inhibition and neurotoxicity. *Journal of Biological Chemistry 285*, 36070-36080.

103. Thomas, S. R., McTamney, P. M., Adler, J. M., LaRonde-LeBlanc, N., and Rokita, S. E. (2009) Crystal structure of iodotyrosine deiodinase, a novel flavoprotein responsible for iodide salvage in thyroid glands. *Journal of Biological Chemistry 284*, 19659-19667.

104. Preiss, T., Belin, S., Beghin, A., Solano-Gonzàlez, E., Bezin, L., Brunet-Manquat, S., Textoris, J., Prats, A.-C., Mertani, H. C., Dumontet, C., and Diaz, J.-J. (2009) Dysregulation of ribosome biogenesis and translational capacity is associated with tumor progression of human breast cancer cells. *PLoS ONE 4*, e7147.

105. Gnidehou, S., Caillou, B., Talbot, M., Ohayon, R., Kaniewski, J., Noel-Hudson, M.-S., Morand, S., Agnangji, D., Sezan, A., Courtin, F., Virion, A., and Dupuy, C. (2004) Iodotyrosine dehalogenase 1 (dehal1) is a transmembrane protein involved in the recycling of iodide close to the thyroglobulin iodination site. *The FASEB Journal*, 04-2023fje.

106. Moreno, J. C., Klootwijk, W., van Toor, H., Pinto, G., D'Alessandro, M., L√®ger, A. n., Goudie, D., Polak, M., Grv°ters, A., and Visser, T. J. (2008) Mutations in the iodotyrosine deiodinase gene and hypothyroidism. *New England Journal of Medicine 358*, 1811-1818.

107. Dai, G., Levy, O., and Carrasco, N. (1996) Cloning and characterization of the thyroid iodide transporter. *Nature 379*, 458-460.

108. Dohan, O., De la Vieja, A., Paroder, V., Riedel, C., Artani, M., Reed, M., Ginter, C. S., and Carrasco, N. (2003) The sodium/iodide symporter (nis): Characterization, regulation, and medical significance. *Endocrine Reviews 24*, 48-77.

109. Vaidya, B., and Pearce, S. H. S. (2008) Management of hypothyroidism in adults. *BMJ 337*.

110. Pape, T., and Schneider, T. R. (2004) Hkl2map: A graphical user interface for macromolecular phasing with shelx programs. *Journal of Applied Crystallography 37*, 843-844.

111. Sheldrick, G. (2008) A short history of shelx. *Acta Crystallographica Section A 64*, 112-122.

112. Sheldrick, G. M. (2002) Macromolecular phasing with shelxe. *Zeitschrift f√°r Kristallographie 217*, 644-650.

113. Watson, J. A. (2006) Insight into the structure and mechanism of iodotyrosine deiodinase, the first mammalian member of the nadh oxidase/ flavin reductase

superfamily. In *Chemistry and Biochemistry Theses and Dissertations* University of Maryland, College Park. http://drum.lib.umd.edu/handle/1903/3794

114. Tanner, J. J., Lei, B., Tu, S.-C., and Krause, K. L. (1996) Flavin reductase P: Structure of a dimeric enzyme that reduces flavin. *Biochemistry 35*, 13531-13539.

115. Koike, H., Sasaki, H., Kobori, T., Zenno, S., Saigo, K. Murphy, M.E.P., Adman, E.T., Tanokura, M. (1998) 1.8 Å crystal structure of the major NADPH:FMN oxidoreductase of a bioluminescent bacterium, *Vibro fisheri*: Overall structure, cofactor and substrate-analog binding, and comparison with related flavoproteins. *Journal of Molecular Biology 280*, 259-273.

116. Taga, M. E., Larsen, N. A., Howard-Jones, A. R., Walsh, C. T., and Walker, G. C. (2007) Blub cannibalizes flavin to form the lower ligand of vitamin B12. *Nature 446*, 449-453.

117. Holm, L., and Park, J. (2000) Dalilite workbench for protein structure comparison. *Bioinformatics 16*, 566-567.

Seth R. Thomas

3400 Pennsylvania St. Hyattsville, MD 20783
937-623-3653
sethomas101@gmail.com

EDUCATION

Ph.D in Biochemistry. 2006-Current
University of Maryland College Park. Department of Chemistry and Biochemistry
Doctoral Graduate Student, Dr. Nicole LaRonde-LeBlanc

B.S. in Biochemistry. 2006
The Ohio State University
Bachelors of Science Biochemistry
Minor: General Business Administration

PROFESSIONAL EXPERIENCE

University of Maryland College Park, 2006-Current
Graduate Research Assistant/Teaching Assistant

- **Research Experience**: My research focus at the University of Maryland has been on structural studies of ribosome processing factors. During my tenure, I have worked on an atypical Rio kinase (Riok3) and Emg1, an N1-pseudouridine RNA methyltransferase. We have shown primary sequence and secondary structure specificity through a crystal structure of Emg1 bound in complex as an obligate dimer with target RNA, containing a base-flipped uridine (substrate analog of pseudouridine to promote crystallization) and a hairpin loop spanning across the dimer interface. I have also collaborated on multiple projects including a breakthrough on a novel structure of iodotyrosine deiodinase (IYD), which salvages iodide from mono- and diiodotyrosine via a dehalogenation mechanism. Without IYD's salvage pathway in thyroid glands, iodide deficiency can lead to hypothyroidism. The combination of diverse projects has given me the experience needed to pursue a post-graduate research position using what I've learned in project design, cloning, protein purification, crystallization, data collection (both home source and synchrotron), and structure solution/refinement.
- Organic laboratory teaching assistant
- General biochemistry teaching assistant

JamesCare for Life, Arther G. James Cancer Hospital and Richard J. Solove Research Institute Columbus, OH. 2005-2008
Student Administrative/Research Assistant

- Keep team members informed of current cancer research throughout the world.
- Accelerate responses to research/grant opportunities.
- Benchmark and develop patient support events.

The Ohio State University Columbus, OH. 2003-2004
Resident Advisor

- Led 48 students throughout their first year experience

- Facilitated floor meetings and planned/executed floor programs
- Actively Involved in Biological Sciences Living Learning Program

TECHNICAL EXPERIENCE

- Structural Biology – Extensive training in X-ray crystallography including growing/optimizing/freezing crystals, collecting data using the synchrotron and in-house sources (Bruker), data processing (HKL2000, Proteum) and structure refinement (Phenix, CCP4, COOT), experience using both molecular replacement and single wavelength anomalous dispersion (SAD) for phase determination (PHASER, SHELX, Solve/Resolve, ARP/wARP) using both sulfur and selenium anomalous diffraction.
- Biochemistry – Radioactive labeling/purification of RNA for footprinting/sequencing/nitrocellulose/DEAE binding assays, fluorescence anisotropy using labeled RNA, thermal shift assay using real-time PCR, analytical ultracentrifugation (sedimentation), mass spectrometry, gel-shift assays, Protein/RNA complex crystallization.
- Protein Purification –Affinity chromatography, gel filtration, ion exchange, widespread experience troubleshooting/protocol optimization of purification protocols.
- Molecular Biology – Recombinant cloning, PCR, DNA and RNA purification, sequencing, transformations, protein expression, site-directed mutagenesis, mammalian cell culture
- Lab Management –Maintaining and repairing Bruker x-ray diffraction system, maintain and repair ALL laboratory equipment, including Phoenix high-throughput liquid handler and chromatography systems, laboratory inventory and ordering.

PUBLICATIONS

1. Thomas, S. R., McTamney, P. M., Adler, J. M., LaRonde-LeBlanc, N., and Rokita, S. E. (2009) Crystal Structure of Iodotyrosine Deiodinase, a Novel Flavoprotein Responsible for Iodide Salvage in Thyroid Glands. *Journal of Biological Chemistry.* 284. 19659-19667.
2. Thomas, S. R., Keller, C., Szyk, A. Cannon, J. LaRonde-LeBanc, N. (2010) Structural Insight into the Functional Mechanism of Nep1/Emg1 N1-specific Pseudouridine Methyltransferase in Ribosome Biogenesis. *Nucleic Acids Research.* Advance online publication.

PROTEIN DATA BANK SUBMISSIONS

1. 2009 - 3GB5 *M. musculus* IYD bound to MN
2. 2009 - 3GFD *M. musculus* IYD bound to FMN and MIT
3. 2009 - 3GH8 *M. musculus* IYD bound to FMN and DIT
4. 2010 - 3O7B *A. fulgidus* Nep1 bound to S-adenosylhomocysteine (SAH)
5. 2010 - 3OII *S. cerevisiae* Emg1 bound to SAH
6. 2010 - 3OIN *S. cerevisiae* Emg1 bound to SAH and 1 molecule of RNA
7. 2010 - 3OIJ S. cerevisiae Emg1 bound to SAH and 2 molecules of RNA

AWARDS/FELLOWSHIPS

- Achievement Awards for College Scientists (A.R.C.S.) Foundation. 2009-2010 and 2010-2011 Metro Washington fellowship recipient.
- Jacob K. Goldhaber International Travel Award. American Crystallographic Association 2009 Toronto Conference. Poster presentation. "Crystal Structure of Iodotyrosine Deiodinase, a Novel Flavoprotein Responsible for Iodide Salvage in Thyroid Glands."
- "2005 Making a Difference Award." Presented to JamesCare for Life team for Outstanding Customer Service. October 2005.

RESEARCH PRESENTATIONS

- July 2009. American Crystallographic Association International Conference. Toronto, Ontario, Canada. Poster presentation.
- May 2009. 39th Macromolecular Mid-Atlantic Crystallography Conference. University of Maryland College Park. Poster presentation
- April 2009. A.R.C.S. Seminar and Presidential Luncheon. University of Maryland College Park. Oral presentation.
- November 2008. Bioscience Day. University of Maryland. Poster presentation.
- June 2007. 37th Macromolecular Mid-Atlantic Crystallography Conference. University of Virginia. Poster presentation.

REFERENCES

Dr. Nicole LaRonde-LeBlanc (Assistant Professor)
University of Maryland, Department of Chemistry and Biochemisty
Biomolecular Sciences Bldg #296
College Park, MD 20742
301-405-0462
nlaronde@umd.edu

Dr George H. Lorimer (Distinquished University Professor)
Director of Center for Biomolecular Structure and Organization
University of Maryland, Department of Chemistry and Biochemistry
Biomolecular Sciences Bldg #296
College Park, MD 20742
301-405-1828
glorimer@umd.edu

Dr. Dorothy Beckett (Professor)
University of Maryland, Department of Chemistry and Biochemistry
Biochemistry Building #091
College Park, MD 20742
301-405-1812
dbeckett@umd.edu