Spring 5-2017

Uncovering the Identity and Metabolism of Bacterial CoA-RNA

Joseph R. Spangler
University of Southern Mississippi

UNCOVERING THE IDENTITY AND METABOLISM OF BACTERIA COA-RNA

by

Joseph Rankin Spangler

A Dissertation
Submitted to the Graduate School
and the Department of Chemistry and Biochemistry
at The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

Approved:

_____
Dr. Faqing Huang, Committee Chair
Professor, Chemistry and Biochemistry

_____
Dr. Sabine Heinhorst, Committee Member
Professor, Chemistry and Biochemistry

_____
Dr. Vijayaraghavan Rangachari, Committee Member
Associate Professor, Chemistry and Biochemistry

_____
Dr. Douglas Masterson, Committee Member
Associate Professor, Chemistry and Biochemistry

_____
Dr. YanLin Guo, Committee Member
Professor, Biological Sciences

_____
Dr. Karen S. Coats
Dean of the Graduate School

May 2017

Published by the Graduate School

THE UNIVERSITY OF SOUTHERN MISSISSIPPI®

ABSTRACT

UNCOVERING THE IDENTITY AND METABOLISM

OF BACTERIA COA-RNA

by Joseph Rankin Spangler

May 2017

Coenzyme A is an indispensable molecule in all known life with roles in metabolism, gene regulation, and macromolecule synthesis. As CoA is derived from RNA itself, it's incorporation into RNA by *in vitro* methods has proven useful in research probing the origin of life based on the RNA World theory. The discovery in contemporary bacteria of RNA modified with CoA, however, provided an unexpected twist to previously well-characterized bacterial systems. The identity of sequences associated with CoA-RNA has been elusive since their discovery in 2009 based on the difficulties in isolation while maintaining RNA quality. The aim of this study is to develop a specific CoA-RNA capture protocol to uncover the sequence identities for the further understanding of CoA-RNA function. While the free thiol of CoA can be used to purify the CoA-RNA from total bacterial RNA, the success of process depends on the specific release from the purification matrix. This work details the characterization of recombinant Nudix CoA pyrophosphohydrolases from various organisms for their use in CoA hydrolysis and CoA-RNA metabolism, as well as for the development and successful implementation of a specific CoA-RNA capture protocol. The use of this protocol for the construction of high quality NGS libraries is vital to the elucidation of CoA-RNA function via bioinformatics analysis, for its potential

ranges from a simple stabilization adduct to contemporary examples of RNA World era coenzyme-conjugated ribozymes.

ACKNOWLEDGMENTS

This work was made possible by the ongoing communication and openness of the Department of Chemistry and Biochemistry, especially the Rangachari, Heinhorst, and Guo labs. I would like to convey special thanks to symbolize my deep gratitude to my advisor, Dr. Faqing Huang, whose patience was as unending as his inspirational drive to obtain the best results from any situation.

DEDICATION

This work is dedicated to my wife and family, whose patience and support were unquestioning and absolute.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF SCHEMES

LIST OF ABBREVIATIONS

ATP, pppA	Adenosine triphosphate
ADP, ppA	Adenosine diphosphate
AMP, pA	Adenosine monophosphate
3',5'-ADP	3', 5'-adenosine diphosphate
pAp	3', 5'-adenosine diphosphate
PhP	4'-phosphopantetheine
NaAc	Sodium Acetate
CoA	Coenzyme A
depCoA	dephospho CoA
CoAse	CoA pyrophosphohydrolase
NAD	Nicotinamide adenine dinucleotide
NMN	Nicotinamide mononucleotide
FAD	Flavin adenine dinucleotide
FMN	Flavin mononucleotide
SAM	*(S)*-adenosyl methionine
ACP	Acyl carrier protein
LUCA	Last universal common ancestor
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
RNase	Ribonuclease
RNasin	RNase Inhibitor
RNAP	RNA polymerase

DTT	Dithiothreitol
rRNA	ribosomal RNA
tRNA	transfer RNA
mRNA	messenger RNA
sRNA	small non-coding RNA
siRNA	small interfering RNA
miRNA	Micro RNA
piRNA	Piwi-interacting RNA
ncRNA	non-coding RNA
pppRNA	5’ triphosphate RNA
pRNA	5’ monophosphate RNA
RNAcho	2’,3’-dialdehyde RNA
IVT	In Vitro Transcription
PCR	Polymerase Chain Reaction
RT	Reverse Transcription
RTase	Reverse Transcriptase
UTR	untranslated region
ORF	open reading frame
dsRNA	double-stranded RNA
ssRNA	single-stranded RNA
nt	nucleotide
bp	basepair
Da	Dalton

kD	kilodalton
NGS	Next Generation Sequencing
TS-RT	Template-switch reverse transcription
RBS	ribosomal binding site
Pi	inorganic phosphate
PPi	*pyrophosphate*
dGTP	deoxyguanosine triphosphate
dATP	deoxyadenosine triphosphate
dNTP	deoxynucleotide triphosphate
NTP	nucleotide triphosphate
7mG	7-methyl guanosine
AppA	Diadenosine diphosphate
s4U	4-thiouridine
s2U	2-thiouridine
MBP	maltose binding protein
TrxA	thioredoxin
LB	Luria broth
Kan	kanamycin
Amp	ampicillin
Cam	chloramphenicol
SDS	sodium dodecyl sulfate
PAGE	polyacrylamide gel electrophoresis
HPLC	high pressure liquid chromatography

SAX	strong anion exchange
UV	ultraviolet light
T4PNK	T4 Polynucleotide Kinase

CHAPTER I – INTRODUCTION

1.1 RNA

1.1.1 Origin of Life

The greatest question science has sought to answer is that of the origin of life. With all of the diversity and complexity of biological systems on the Earth, many have turned their sights backwards in contemplation of where life began. The outright answer to this question may never be known due to the loss of direct scientific evidence over the course of the planet's 4.5 billion year history. We therefore are left with the task of theorizing with the evidence that is available. As time progresses and new evidence is uncovered, theories are amended in light of the new perspectives given by that evidence (Joyce 1989). As a result, some theories become stronger while others become obsolete.

An example of a strong theory regarding origin with respect to contemporary organisms is that of a common ancestor. As evaluation and comparison of highly conserved RNA sequences points out, there is a branching point from which the three kingdoms of life diverged known as the Last Universal Common Ancestor (LUCA). Regardless of the shifting timelines of organismal divergence that change with different phylogenetic analyses, one remaining constant is that of the LUCA. Due to the overwhelming similarities among the kingdoms of essential biochemical processes such as DNA replication, RNA transcription, protein translation, and basic metabolism, their emergence from a common ancestor is a highly accepted theory. However, the LUCA itself is a

theoretical organism that cannot be studied, and therefore conclusions regarding this must be made with reservation and only with great supporting evidence.

RNA is an important molecule regarding evolution and the origin of life. This fact was recognized by Carl Woese when he used comparisons of 16S ribosomal RNA (rRNA) to construct the phylogenetic tree identifying Archaea as a kingdom different from Eubacteria. The identification that RNA is ancient enough to differentiate all known organisms implicates its existence at the dawn of our contemporary biological world. This point is further supported by the essential role of RNA in the indispensable process of protein translation. A quick evaluation of the logic that RNA is required for protein synthesis as a template (messenger RNA), as a carrier molecule for monomers (transfer RNA), and as the catalytic molecule to form the peptide bond (ribosomal RNA) allows one to hypothesize that RNA predates the evolution of protein. Similarly, as RNA is essential to DNA synthesis both as a primer for DNA replication and as a substrate for ribonucleotide reductase to produce deoxyribonucleotides, it can be hypothesized that RNA also predates the evolution of DNA. The integral role in life played by RNA, the capability for informational storage, and the catalytic properties of RNA are all ideas that compose the theory of the RNA World (Gilbert 1986).

The RNA World is an intriguing idea wherein RNA was the principle molecule supporting life before the advent of DNA and protein. One intrinsic problem with this theory however is the lack of modern day examples of RNA molecules that would support life by a complex metabolism and self-replication.

These processes today are largely carried out by protein enzymes with such high efficiency and specificity it's hard to comprehend a chaotic existence without them. The RNA World however is not meant to be convincing of the superiority of RNA function over that of proteins, but rather of such functions as a vestige of a previous era (Benner et al. 1989, White 1982). Support of the RNA World theory relies on examples RNA-catalyzed fundamental processes that would be required in an RNA organism, such as RNA-catalyzed formation of thioester bonds (Coleman and Huang 2002), acyl bonds (Li and Huang 2005), peptide bonds, phosphodiester bonds (Yarus 1999), and coenzyme synthesis (Huang et al. 2000). While these do not paint the entire picture of an RNA-centered existence, they provide evidence that fundamental chemistries can be carried out by RNA alone. Consideration of these discoveries as a precursor to facilitating more complex processes such as template-derived polymerization of nucleotide triphosphates, energy storage via bond formation and hydrolysis, and improvement of catalytic efficiency by coenzymes infers the importance of RNA to the evolution of the modern organism.

Considering that the RNA World can be reasoned to predate the LUCA (Benner et al. 1989), the discovery of RNA-catalyzed processes to support the theory relies heavily on *in vitro* methods for proof-of-concept. Only time will tell whether more evidence arises in support of or conflict with the RNA World theory. The proven ability of RNA to carry genetic information, catalyze important biochemical reactions, and fold in three-dimensional structures to potentially bind any molecule certainly implicate RNA to be capable of doing everything protein

can, and perhaps more. Whether or not an RNA organism is the predecessor to the complex biological systems of today, there is no denying the current functional diversity of RNA and its impact on life as we know it.

1.1.2 Contemporary RNA Functions

RNA has been considered a central part of life since the idea of the central dogma was introduced. The role of messenger RNA for example is integral to survival, for it is the obligate intermediate to protein synthesis required by all living organisms. High fidelity mRNA production is essential to protein function, as the sequence elements of mRNA determine the rate of translation and the amino acid composition of a protein, and therefore the abundance and primary structure of a protein. Furthermore, mRNA sequence elements determine the alternative splicing in higher organisms which impacts protein diversity. Although mRNA is the recognized molecule in the central dogma, the majority of RNA in a cell (~90%) is tRNA and rRNA (Zhuang et al. 2012). The defined structure of tRNA in concert with the myriad nucleotide modifications it carries is responsible for the placement of the correct amino acid encoded by the mRNA into the ribosome during translation. The ribosome itself is a complex of protein and rRNA, where the catalytic activity of peptide bond formation is carried out by the rRNA component. The functions of these three major classes of RNA are well known in the scientific community, and the removal of one of these would certainly be deleterious to an organism. These three distinct classes of RNA with their conserved structures and functions only comprise part of the role of RNA in a cell.

New discoveries of RNA functions have added insight into the diversity and importance of RNA in a cell. Aside from the active roles played in translation by mRNA, tRNA, and rRNA, there are other classes that recently have been discovered. These RNA are considered non-coding RNA (ncRNA), and comprise a large variety of RNA involved in one way or another with regulation of gene expression. The ncRNAs are found in both prokaryotes and eukaryotes and serve to directly interact with mRNA, DNA, or proteins to alter gene expression. Examples of eukaryotic ncRNA are microRNA (miRNA), small-interfering RNA (siRNA), and piwi-interacting RNA (piRNA).

MicroRNA represents a class of small RNA involved in translation regulation. The impact of miRNAs is astounding in eukaryotes, for their origins are in intron sequences of DNA that previously had an unknown function and were largely thought to be “junk”. MicroRNA regulates protein expression by interacting with 3’ untranslated regions (3’UTRs) of mRNA. Over 60% of human mRNA contain miRNA targeting sequences, and there are over 2500 catalogued miRNA (miRBase.org) indicating that miRNA are a very active regulatory mechanism (Ha and Kim 2014). Derived from introns, primary miRNA is transcribed by RNA Polymerase II (PolII) to >1kb single-stranded RNA that contains a stem-loop structure. That structure contains the miRNA sequence that is released or matured by RNase III enzyme Drosha to form pre-miRNA, which is then further matured by the enzyme DICER to 22 basepair (bp) double-stranded RNA (dsRNA) that is an active miRNA (Ha and Kim 2014). The multiple rounds of processing occur by different enzymes and in different portions of the cell,

implying this miRNA pathway to be the resulted of a concerted effort of proteins and transport. Once the mature miRNA is produced, it associates with AGO proteins to form the RNA-induced silencing complex (RISC) that is responsible for cleavage of mRNA targeted by complementary base pairing of 7 nucleotides within the guide miRNA seed sequence (Ha and Kim 2014). The small number of base pairs required for AGO-miRNA activity gives miRNA a broad targeting range of mRNA, and is therefore very important to its function. MicroRNA function, if perturbed, can result in tumor formation due to deregulation of protein expression, and therefore miRNA currently serve as biomarkers used in cancer research (Dong et al. 2013).

Small interfering RNA (siRNA) is another small RNA in eukaryotes that associates with AGO proteins. Similar to miRNA, siRNA association with AGO proteins mediates the cleavage of RNA based on complementary base pairing. While the end function is the same with the exception of siRNA acting as the targeting guide rather than miRNA, their origins are quite different. Small interfering RNA is derived from viral RNA and is used as an antiviral defense. Viral infection of cells is consists of abundant transcription of viral mRNA and reproduction of the viral genome encoding these mRNA. Long dsRNA that results from such reproduction of viral genomic RNA then serves as a hallmark for viral infection, and these long dsRNA serve as a substrate for processing by the Dicer enzyme into siRNA (Wilson and Doudna 2013). The siRNA then can be used to selectively target and degrade their viral counterparts resulting in the ablation or knockdown of gene expression. This process of RNA interference

(RNAi) has been exploited since its discovery as an alternative to gene knockouts to study gene function in eukaryotes, where knowledge of a gene-of-interest can be used to design dsRNA of 25 nucleotides (nts) and transfected into cell lines for the observed RNAi effect of temporary gene knockdown.

Piwi-interacting RNA (piRNA) represent another class of small RNA that interacts with AGO proteins. More specifically, piRNA interact with the Piwi clade of the Argonaute superfamily. Piwi-interacting RNA are 24-30nt in length and are active in germ cell lines and during development (Aravin et al. 2007). Their function has not completely been characterized, but some studies have shown piRNA to protect genomes against parasitic nucleic acid such as transposons. These RNAs are not conserved like miRNA, and therefore their conserved functions seem to be elusive. Conceptually, however, they appear similar to their small RNA counterparts of siRNA and miRNA in that there is an association with nucleotide-hydrolyzing proteins and thus a function in regulation or protection of some sort. These small regulatory RNAs are relatively new to the scientific community, and there is still much to be discovered about them.

Prokaryotes do not specifically contain miRNA, siRNA, or piRNA, but they are not without RNA-mediated post-transcriptional regulation. Small noncoding RNAs (sRNA) in bacteria can be considered a prokaryotic version of post-transcriptional gene regulation akin to miRNA in eukaryotes. These bacterial sRNA are 50-300nt in size with a 7-12 bp seed region similar to miRNA (Bandyra et al. 2012). While miRNA form the RISC complex with Dicer, sRNA form a complex with Hfq proteins that will associate with the RNase E degradasome to

facilitate the degradation of the target mRNA (Bandyra et al. 2012). Antisense gene regulation in bacteria has been known since the 1980s, wherein RNA binding is followed by negative regulation either by sequestration of primers for DNA replication, or binding mRNA to block protein translation (Wagner and Simons 1994). The more recent discovery of the RNA-protein complex mediating degradation of target mRNA makes the system more analogous to miRNA in eukaryotes. The association of Hfq with these sRNA also implicates the regulatory role of sRNA in bacterial virulence.(Papenfort and Vogel 2010) The full details of sRNA regulation, especially with Hfq association, are yet to be determined. Recent advances in Next Generation Sequencing (NGS) technology have allowed investigators to predict more targets for known sRNA, and have also identified new potential sRNA.(Cahova et al. 2015, Papenfort and Vogel 2010) In addition to specific mRNA regulation, conserved Hfq-dependent sRNA have been implicated in the activation of σ^s synthesis (Papenfort and Vogel 2010). Bacteria contain a single RNA polymerase responsible for all transcription, and therefore the determining factors of transcript abundance are promoter sequence and σ-factor participation in transcription initiation. The sRNA control of which σ-factors are expressed exemplifies yet another level of transcriptional control exerted by sRNA. The field of sRNA regulation in bacteria is sure to continue to grow as the cost of NGS continues to reduce and become more feasible for incorporation into experimental design.

The role of RNA in gene regulation was recently expanded with the discovery of Clustered, Regularly Interspaced Short Palindromic Repeats

(CRISPR) in bacteria and archaea in recent years. These RNA sequences were discovered to be analogous to an adaptive immunity in bacteria capable of translational repression by targeted degradation of mRNA. Sequences are acquired from a primary infection by bacteriophage or plasmid and incorporated into the genome as repeated units separated by spacers (Marraffini and Sontheimer 2010). Infections from different sources provide the breadth of the repeat units, and upon secondary infection the transcription of these CRISPR RNA (crRNA) sequences facilitates the double-strand breakage of the cognate DNA to combat the infection. Mechanistic investigations uncovered the potential of the Type II CRISPR-Cas9 system for genome editing (Jinek et al. 2012). The Type II system requires trans-activating crRNA (tracrRNA) to basepair with crRNA to induce processing by RNase III for activation (Jinek et al. 2012). While in the cell these two RNA are transcribed separately, they can be ectopically transcribed as a single molecule with the same structure to elicit the same response. Therefore by designing a tracrRNA:crRNA to target a portion of the genome, one can induce a double-strand breakage at that site. Furthermore, by providing a homologous replacement source of that site in the genome, the homologous recombination repair mechanisms of the bacteria can effectively add a synthetic sequence to the targeted region of the genome resulting in point mutation or gene deletions (Jinek et al. 2012). The process of long pre-crRNA transcription followed by protein-mediated maturation into a shorter active form makes CRISPR analogous to the eukaryotic miRNA pathway (Marraffini and Sontheimer 2010). The idea of an RNA-driven adaptive immunity in prokaryotes

further displays the variety of RNA functions that exist. While higher diverged organisms such as mammals display a more complex and diverse adaptive immunity dependent on protein, the existence of the CRISPR system in simpler prokaryotes supports the theory of RNA-stabilized processes that throughout evolution were amended with protein to more complex and diverse processes.

Bacterial gene regulation can also occur in a *cis*-acting manner. Transcribed mRNA must have an accessible promoter and RBS for translation initiation, and these sequence elements appear upstream of the open reading frame (ORF) in the 5'UTR. Riboswitches are mRNAs that contain an aptamer domain in the 5'UTR that will bind a metabolite with high affinity and induce structural changes to allow or block translation of the encoded gene (Tucker and Breaker 2005). This action allows the tunable translation of certain genes based on the cellular environment. In addition to metabolite-binding riboswitches, bacterial mRNA can also possess similar mechanisms that act as thermosensors. The thermosensor is a structural element in the 5'UTR that will change conformation based on temperature (Papenfort and Vogel 2010). In lower temperatures, for example, RNA secondary structure may form blocking the RBS, and rising temperature denatures that structure to expose the RBS and make the mRNA accessible for translation. Similar mechanisms can occur with pH changes (Papenfort and Vogel 2010). Riboswitches are abundant in bacteria, but are unknown in eukaryotes with the exception of one in plants (Bocobza and Aharoni 2008, Tucker and Breaker 2005). Without representation of the riboswitch in more organisms, the evolutionary origin of such a mechanism is

uncertain. Regardless, the riboswitch mechanism is interesting as it involves a single self-regulating RNA with a structure capable of high affinity binding and a gene-encoding portion, and itself is a great example of the diverse multi-functional nature of RNA.

The ribozyme represents a different aspect of RNA function from those previously discussed. The roles of RNA in post-transcriptional regulation and antiviral protection are relatively passive roles, where the targeting is carried out by RNA but the catalytic mechanism is carried out by protein. In the case of ribozymes, the RNA is responsible for both the targeting and the catalytic mechanism. The activity of ribozymes, much like protein enzymes, is reliant on the secondary structure of the RNA which is in turn depends on the sequence itself or primary structure. Binding is made possible by taking on specific folding patterns that specifically orient substrates and make certain chemical reactions favorable. Naturally occurring ribozymes are quite important for survival, such as the activity of small nuclear RNA (snRNA) to mature eukaryotic mRNA by removing introns. The removal of these introns and joining of exons serves to create protein isoforms in a splicing process that is carried out by the majority of human protein-encoding genes (Dvinge et al. 2016). It is widely accepted that splicing accounts for the abundance of diverse protein functions encoded by a relatively small percentage of the human genome. The splicing activity of snRNA on a conceptual level is the same as small-scale polymerization, or the formation of a phosphodiester bond. Ribozymes derived from *in vitro* selection methods catalyze a diverse set of reactions that have not yet been identified in natural

systems. In addition to phosphodiester bond formation carried out by snRNA, *in vitro* selected ribozymes have been found that carry out phosphodiester formation to act as a 5'-3' and a 5'-5' ligase, as well as phosphodiester hydrolysis and even RNA branching and polymerization to some scale (Tuschl et al. 1998, Pan and Uhlenbeck 1992, Bartel and Szostak 1993, Chapman and Szostak 1995, Ekland and Bartel 1996). In addition to the 5'-5' ligase activity shown by Chapman et al. (1995), 5'-5' self-capping ribozyme activity has also been observed (Huang and Yarus 1997). The formation and hydrolysis of phosphodiester bonds by RNA is certainly impressive, but investigators have observed much more diverse ribozyme-catalyzed chemical reactions such as RNA phosphorylation, acyl activation, aminoacylation, acyl transfer, amide bond formation, glycosidic bond formation, RNA alkylation, thioester formation, Michael addition, Diels-Alder reactions, and redox reactions (Lorsch and Szostak 1994, Kumar and Yarus 2001, Illangasekare et al. 1995, Lohse and Szostak 1996, Unrau and Bartel 1998, Wilson and Szostak 1995, Jadhav and Yarus 2002, Coleman and Huang 2002, Sengle et al. 2001, Tarasow et al. 1997, Tsukiji et al. 2003). The list of ribozyme-catalyzed chemistries is impressive, but these activities are evidence of the potential for RNA to support life itself. While the majority of ribozyme chemistries that have been observed involve a phosphate or an anhydride, the discovery of ribozymes catalyzing carbon-carbon bond formation, glycosidic bond formation, and alcohol dehydration broadens the potential for ribozyme activity to be capable of supporting life itself (Chen et al. 2007).

The importance of RNA in life has been argued for many years (Gilbert 1986). It is accepted as an obligate part of every organism by the central dogma of biology. While serving as the template for protein translation is indispensable for cell survival, it's clear the role of RNA extends far beyond that of an intermediate. RNA has been shown to serve as an information carrier, regulatory, and catalytic molecule in all forms of life. In terms of catalytic RNA, the diverse array of ribozyme activity observed both *in vitro* and in nature is astounding. Although most ribozyme discoveries are via *in vitro* selection, the resulting activity is evidence of the potential within RNA, and evidence that such ribozymes may have existed in a time when the selective pressure was present before the emergence of protein. It seems with all that has been recently discovered, the depth and scope of RNA function has yet to be fully uncovered.

1.2 Coenzymes

1.2.1 Origin and Functions

Coenzymes are molecules that participate in reactions with protein enzymes, and are considered the most conserved complex molecule throughout evolution (White 1982). In the 1970s, half of all known enzymes required coenzymes for their function (White 1976). Today, such coenzymes are ubiquitous throughout life and indispensable for protein function in important pathways such as glycolysis and the Kreb's cycle. Many coenzymes contain modified nucleotides or derivatives such as coenzyme A (CoA), nicotinamide adenine dinucleotide (NAD), Flavin adenine dinucleotide (FAD), S-adenosyl methionine (SAM), thiamin pyrophosphate, and pyridoxal phosphate (Figure 1.1).

The housing of nucleotide structures at the conserved core of coenzymes understood to be vital to modern enzyme function suggests a time when nucleotide-derived enzymes, or ribozymes, fulfilled such roles (White 1976, White 1982). Interestingly, the nucleotide incorporation plays no role in the function of these coenzymes, and therefore is considered a vestige of a previous era when these coenzymes likely existed as the catalytic portion of ribozymes (Benner et al. 1989). Due to the ability for coenzymes to carry out the same reactions as those of proteins that require them, and the known capability for proteins to bind specific substrates and stabilize reaction transition states, it's possible that as evolution progressed, the emergence of proteins granted specificity to the biochemistry catalyzed by ribozymes, and the union of the two resulted in the modern day protein-coenzyme catalytic complex (White 1976, White 1982). Such an evolution would result in the catalytic center being trimmed to the smallest unit still capable of catalysis, resulting in the modern coenzyme with an RNA nucleoside or its derivative at the core.

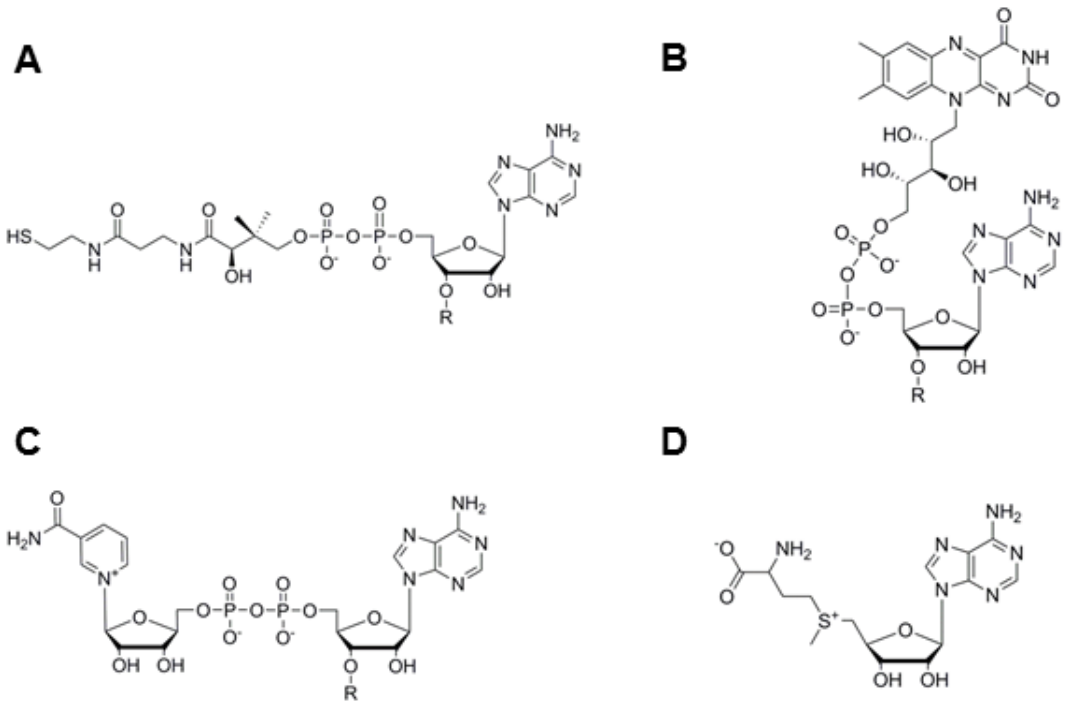


Figure 1.1 Common Coenzymes Derived from Adenosine.

A, Coenzyme A (CoA); B, Flavin adenine dinucleotide (FAD); C, Nicotinamide adenine dinucleotide (NAD+); D, *(S)*-adenosyl methionine (SAM).

The coenzymes of interest here are those that contain adenosine at the core of their structure, such as CoA, NAD, FAD, and SAM (Figure 1.1). The functions of these coenzymes are distinct from one another, even in terms of NAD and FAD that both function in redox. As their origin is hypothesized to be from the LUCA or before, these coenzymes have been deeply integrated into modern metabolism, with each coenzyme participating in a multitude of pathways (Benner et al. 1989). For example, SAM functions as the major methyl donor in the cell that is responsible for, among other things, DNA methylation, and therefore has a profound impact on gene expression (Fontecave et al. 2004, Chiang et al. 1996). SAM has also been shown to donate amino groups,

methylene groups, ribosyl groups, and aminoalkyl groups in different pathways. Another example of a widely used coenzyme is NAD. This coenzyme is widely used in redox reactions by cycling between its oxidized (NAD^+) and reduced (NADH) forms. In aerobic respiration from glucose, NAD^+ is reduced during glycolysis and the Kreb's cycle, and the accumulated NADH is then oxidized in the electron transport chain to drive the major ATP-yielding process of the pathway. In addition to this well-known function of NAD^+/NADH in redox chemistry, NAD can be used for adenylation, ADP-ribosylation in DNA repair and cell signaling through G-coupled-proteins, and as a transcriptional regulator via activation of Sir2p deacetylase activity (Lin 2007, Garten et al. 2009). FAD is another example of a multi-use coenzyme. It also participates in redox reactions similar to NAD, but is a more versatile redox substrate due to its ability to capture either 1 or 2 protons (Walsh 1980). Aside from its well-characterized roles in the Kreb's cycle and the electron transport system, FAD participates in many other reactions with the flavoenzyme class of proteins. Flavoproteins are involved in the oxidation of various metabolites such as amino acids and fatty acids, as well as with protein folding and chromatin remodeling (Joosten and van Berkel 2007, Giancaspero et al. 2013). The nature of the coenzyme dictates its versatile involvement throughout life, as there are many examples to support this.

1.2.2 CoA

Perhaps one of the more important and versatile coenzymes found in all forms of life is Coenzyme A (CoA). While the importance of CoA is open for debate, the structure of CoA is unique and there are no functional overlaps with

other abundant coenzymes such as those with NAD/FAD in redox and SAM/methylcobalamin in methyl transfer. While CoA is known to participate in over 500 reactions, its main function is acyl-transfer (Kanehisa et al. 2008). This activity is made possible by the thiol group of CoA which can react with acyl-phosphates to form thioesters. The thioester serves as a great intermediate for acyl transfer due to its high energy and considerable stability (Coleman and Huang 2002, Berg et al. 2007). Cells take advantage of the acyl-transfer capabilities of CoA most notably in polyketide synthesis, β-oxidation in fatty acid metabolism, and the transition between glycolysis and the Kreb's cycle. CoA also plays a role in protein and amino acid synthesis when considering the intermediates of the Kreb's cycle are required for the synthesis of 10 amino acids (Berg et al. 2007). Recently, CoA has been shown to participate in acetylcholine synthesis and histone acetylation (Szutowicz et al. 2013, Nakamura et al. 2012). These examples place the direct role of CoA in metabolism, protein synthesis, post-translational modification, and gene regulation. The importance of CoA for survival is therefore indisputable.

Coenzyme A synthesis takes place by a conserved pathway involving 5 steps initiated from pantothenate (Figure 1.2). The first step is the phosphorylation of pantothenate to 4'-phosphopantothenate by the enzyme CoaA or pantothenate kinase (PanK), which is the rate-limiting step of the pathway and the focus of feedback inhibition by CoA itself for regulation of CoA synthesis (Leonardi et al. 2005, Jackowski and Rock 1981). After its phosphorylation, 4'-phosphopantothenate is condensed with cysteine by CoaB to

form 4’-phosphopantothenoylcysteine, which is then decarboxylated by CoaC to form 4’-phosphopantetheine. Dephospho-CoA (depCoA) is made in step 4 by the adenylation of 4’-phosphopantetheine (PhP) by CoaD, or phosphopantetheine adenylyl transferase (PPAT), and thereafter depCoA is phosphorylated by CoaE to functional CoA. While the pathway itself is conserved, eukaryotes contain a single dual-function enzyme that catalyzes the same reactions as CoaD and CoaE in plants and bacteria (Leonardi et al. 2005). The CoA pool in eukaryotes also varies from cytoplasm to mitochondria as well as between different tissues because of the different metabolic pathways involved, therefore transcriptional control of PanK is employed to locally regulated CoA levels in addition to feedback inhibition by CoA and its thioesters on PanK activity itself (Leonardi et al. 2005). In *E. coli*, however, the carbon source plays a role in CoA levels, wherein acetate and glucose levels dictate the total amount of CoA as well as the ratio of CoA to thioester (Leonardi et al. 2005). For example, in nutrient rich conditions marked by excess glucose, the CoA levels are high with the majority in the form of acetyl-CoA, but in glucose minimal media lacking pantothenate, the thioester fraction decreases greatly along with the overall CoA concentration (Leonardi and Jackowski 2007).

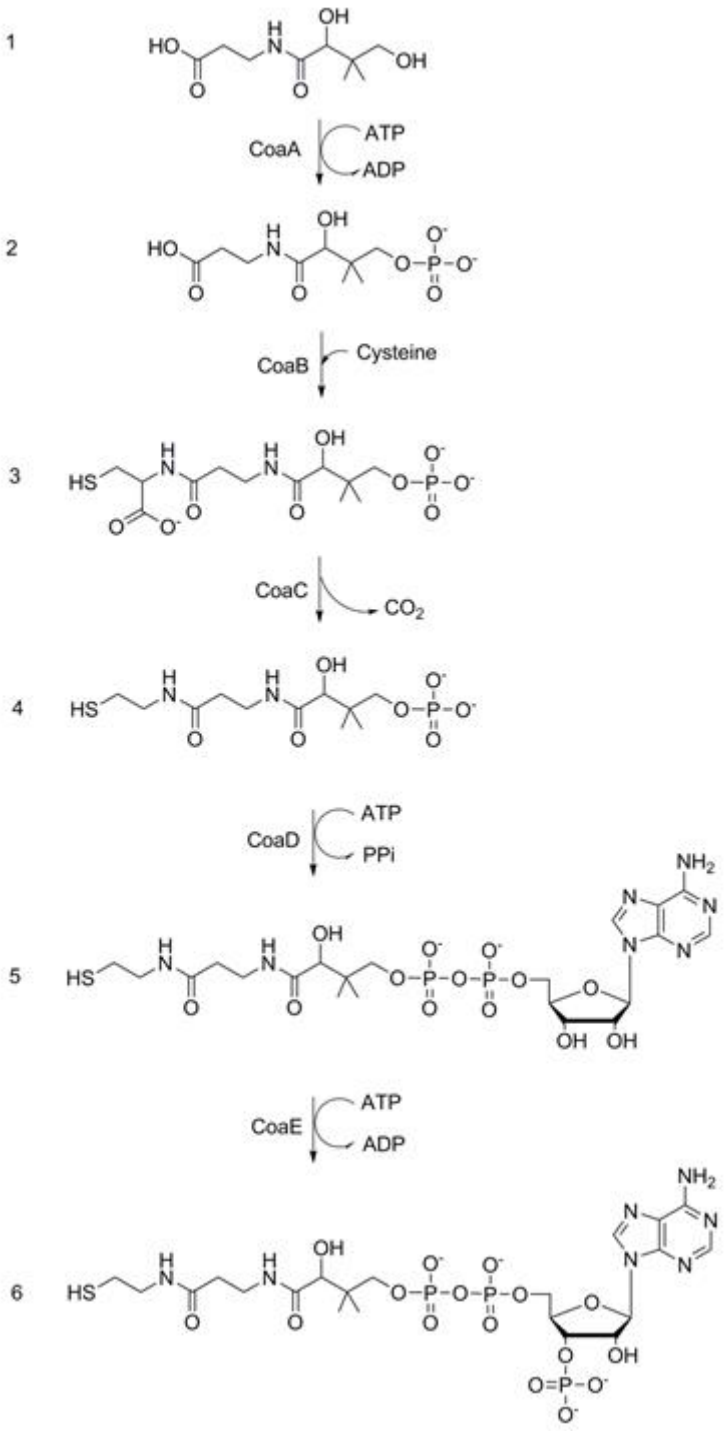


Figure 1.2 Bacterial CoA Synthesis.

Synthesis begins with pantothenate (1) being phosphorylated by CoaA to 4'-phosphopantothenate (2), condensed with cysteine by CoaB to form 4'-phosphopantothenoylcysteine (3), decarboxylated by CoaC to form 4'-phosphopantetheine (4), adenylated by CoaD to make dephospo CoA (5), and phosphorylated by CoaE to form the functional CoA (6).

The source of the CoA synthesis pathway is interesting in terms of evolution. Humans do not possess the ability to synthesize pantothenate, and therefore it must be acquired through other means. *E. coli*, on the other hand, is quite capable of pantothenate synthesis from condensation of the amino acid derivatives pantoic acid and β-alanine (Jackowski and Rock 1981). The interesting aspect of pantothenate synthesis is that it and its precursors can be obtained by prebiotic chemistry. The biosynthesis of pantoate involving methylation and subsequent reduction of α-ketoisovalerate by Methyl-THF and NADPH, respectively, was investigated in 1993 by Miller as a possible prebiotic route to pantoate synthesis, considering that each of the pathway's precursors would be present in the prebiotic ocean (Miller and Schlesinger 1993). In addition to this known pathway, pantoate synthesis was demonstrated with HCN, formaldehyde and the valine precursor isobutraldehyde, wherein the subsequent lactone form of pantoic acid could react with β-alanine to form pantothenate without a catalyst. Both pathways involving isobutraldehyde and α-ketoisovalerate resulted in contextually high yields considering prebiotic reaction conditions (Miller and Schlesinger 1993). Coenzymes have been hypothesized as molecular fossils of evolution since the 1970s, and the demonstration of prebiotic synthesis of coenzyme precursors further supports this claim (White 1976). The existence of CoA as a ubiquitous and vital coenzyme in all of life indicates its existence in the LUCA, and according to Benner the fact that the function of CoA is unrelated to its RNA core implicates its existence before the LUCA in the breakthrough organism during the time of the proposed RNA world

(Benner et al. 1989). Studies in the past 30 years employing *in vitro* selection have yielded a great number of ribozymes as previously discussed, but most notably the Huang lab and the Yarus lab have produced ribozymes that directly interact with CoA (Huang et al. 2000, Coleman and Huang 2002, Li and Huang 2005, Jadhav and Yarus 2002). Such CoA-related ribozyme activity is further evidence of the ancient relationship between CoA and RNA, and supports the theory of an RNA world where coenzymes were synthesized and functioned in a similar way as they are in contemporary life.

The degradation pathway of CoA is peculiarly unknown considering the importance of the molecule. Over the years, CoA synthesis has been a target for antibacterial drug development and therefore the synthesis pathway is fairly well characterized (Leonardi et al. 2005). The transfer of the 4'-phosphopantetheine arm to the acyl carrier protein (ACP) for fatty acid synthesis is also well characterized, but after use, PhP is removed from ACP and recycled to regenerate CoA by the sequential activities of [ACP]phosphodiesterase and PPAT (Jackowski and Rock 1984). The breakdown of depCoA is possible by PPAT which catalyzes both forward and reverse reactions regarding depCoA synthesis and degradation. This degradation cannot occur with CoA itself, however, and requires the dephosphorylation of CoA to depCoA, which is possible by general dephosphorylation with acid phosphatase, but is not characterized by a CoA-specific pathway (Kanehisa et al. 2008, Huang et al. 2000). Alternate pathways of depCoA degradation include ectonucleotide pyrophosphatase (ENPP) activity, but since the 1950s, the only characterized

ENPP enzymes exist in humans (Kanehisa et al. 2008). *E. coli* has been shown to secrete both PhP and pantothenate when in excess, but studies have not uncovered the likewise secretion of CoA, therefore leaving the degradation of CoA to PhP and 3’5’ADP a mystery for many years (Jackowski and Rock 1984, Vallari and Jackowski 1988, Leonardi and Jackowski 2007). If the CoA is not being degraded, then perhaps it is just constantly recycled. Surely, such an important molecule would be maintained in some sort of salvage pathway. The importance of CoA, however, is what makes this aspect of CoA metabolism hard to probe. It is because CoA is such an indispensable component of survival that *coaD* knockout mutants are lethal (Balibar et al. 2011). While knockout mutants can sometimes survive as auxotrophs in supplemented growth media, CoA cannot imported into the cell to rescue *coaD* mutants, therefore making it hard to study CoA retention and recycling by knockout techniques (Balibar et al. 2011). The area of CoA degradation would remain unstudied until the discovery of a new superfamily of nucleotide-hydrolyzing enzymes was found to extend into coenzyme turnover.

1.3 The Nudix Superfamily

1.3.1 Structure and Functions

The degradation of aberrant functioning molecules is a required process in cellular maintenance. Damaging situations can occur as a result of accumulation of such molecules. One such example is the oxidation of the deoxyribonucleotide guanosine (dGTP) to 8-oxo-guanosine (8-oxo-dGTP) by reactive oxygen species in the cell. The oxidized nucleotide can be incorporated into DNA during

replication just as dGTP, but its capability to basepair with adenosine leads to point mutations. In an environment characterized by high oxidative stress, the pool of 8-oxo-dGTP could be large, therefore leading to a large mutation rate during DNA replication. Such was the scenario from which the first Nudix protein, MutT, was discovered in E. coli. (Bessman et al. 1996). The MutT protein was found to cleave nucleotides as a pyrophosphohydrolase, catalyzing the general reaction of dNTP → dNMP + PPi, and due to its preference for oxidized nucleotides such as 8-oxo-dGTP it was determined to be a housecleaning enzyme, responsible for protecting the cell against mutagenic nucleotides (Bessman et al. 1996). Investigation of the amino acid sequence of MutT uncovered a new class of enzymes that catalyze the hydrolysis of nucleotides linked by diphosphate to X (Nudix).

The Nudix box motif (PS00893) appears conserved across archaea, bacteria, eukaryotes, and some viruses (Mildvan et al. 2005). The consensus sequence of Gx_5Ex_5[U/A]xREx_2EExGU (where X is any amino acid, and U is any hydrophobic amino acid) found within a loop-helix-loop motif is the basis for the Nudix superfamily of enzymes, and characterizes the Nudix fold of an αβα sandwich responsible for magnesium/manganese binding and the catalytic activity of the protein (McLennan 2006). Over 250 species have been identified to contain Nudix enzymes, and while there is a general trend of increasing number with genome size, their occurrence is significantly higher in prokaryotes where organisms such as *E. coli* and *D. radiodurans* have 13 and 26, respectively (McLennan 2006, Mildvan et al. 2005). The different Nudix hydrolases found in

bacteria and eukaryotes have been observed to prefer a variety of substrates that satisfy the criteria of a nucleotide diphosphate linked to X, such as 8-oxo-dGTP, 8-OH-dADP, GDP-mannose, GDP-glucose, ADP-ribose, ADP-glucose, ADP-mannose, 5m-dCTP, AppA, Ap_3A, Ap_4A, Ap_5A, diphosphoinositol polyphosphates, 5-phosphoribosyl 1-pyrophosphate (PRPP), thiamin pyrophosphate, NADH, FAD, eukaryotic capped mRNA, and finally CoA and CoA thioesters (McLennan 2006, Bessman et al. 1996, Song et al. 2010). Such a variety is possible, in part, by the substrate ambiguity displayed by many of these enzymes when studied *in vitro* (McLennan 2013). The catalytic activity of all known Nudix enzymes involves hydrolysis of a phosphoanhydride bond with the exception of molecules such as GDP-mannose where the glycosidic bond is targeted. This conserved activity is sufficient to inactivate or prevent the substrate from being used, therefore allowing it to be salvaged for another purpose. The Nudix enzymes can indeed be considered housecleaning enzymes, however their role can be extended into gene regulation which will be discussed in a later section. Furthermore, the activity of Nudix enzymes with coenzymes may indicate their role in the regulation of metabolism by facilitating the turnover of these metabolism-dependent molecules.

1.3.2 Coenzyme Pyrophosphohydrolases

The Nudix enzymes have long been accepted as antimutator proteins and housecleaning proteins in the cell (Bessman et al. 1996). The discovery of an NADH-hydrolyzing Nudix enzyme, however, extended the role of the Nudix superfamily to include the potential for metabolic regulation (Figure 1.4C).

Studies on the NADH-targeting Nudix hydrolases from E. coli, S. cerevisiae, and C. elegans demonstrated a significantly higher preference for reduced NADH than oxidized NAD+, implicating its potential activity to be involved with regulation of the balance between the two (Frick and Bessman 1995, Xu et al. 2000). The importance of the NADH-targeting Nudix enzyme is only speculated, for yeast deletion mutants seem to survive with no significant phenotypic effects (Xu et al. 2000). Some of the speculation involves the production of NMNH (reduced nicotinamide mononucleotide, or reduced nicotinamide D-ribonucleotide), as this is the only known pathway of obtaining such a molecule (Frick and Bessman 1995). Considering the increasing pool of non-canonical functions of nicotinamide coenzymes that has been recently discovered, it is plausible that its role is regulatory. NADH-hydrolyzing Nudix enzymes must maintain some important function, regardless of what is currently known, for the enzyme has been detected in an abundance of organisms based on conserved motif SQPWPxPxS located downstream of the Nudix motif (Xu et al. 2000). The three studied by Xu et al. (2000) from organisms representing prokaryotes, single-celled eukaryotes, and multicellular eukaryotes all display some substrate diversity to various degrees. While showing clear preference for NADH as mentioned earlier, the investigators also observed significantly higher preference for AppA, and ADP-Ribose than for NAD+. The C. elegans enzyme even displayed a higher activity towards dATP than NAD+. Furthermore, the E. coli NudC has recently been shown to cleave NADH, NAD, and CoA, and while the observed substrate promiscuity may indeed be an in vitro bias, it could also

simply be an example of a multi-purpose enzyme that has a preference for NADH (McLennan 2013, Bird et al. 2016).

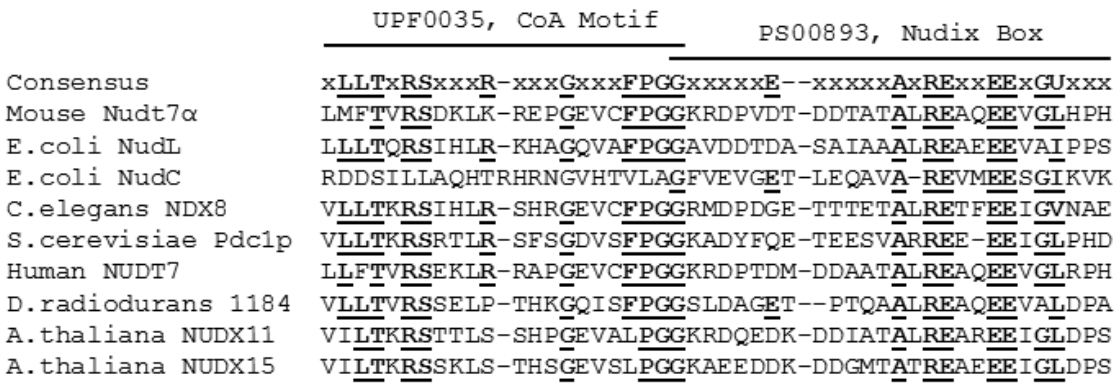


Figure 1.3 Alignment of Nudix Enzymes with CoA Hydrolysis Activity.

Consensus sequences are underlined and emboldened, and all sequences have been verified to have CoAse activity with the exception of E. coli NudL and Human NUDT7. E. coli NudC has no CoA consensus motif, yet still displays CoA hydrolysis activity in vitro.

The CoA-hydrolyzing Nudix enzymes have a similar potential to act as metabolic regulators in addition to their housecleaning capabilities. These enzymes are characterized by a CoA-binding motif of LLTxR[S/A]x_2LRx_3Gx[V/I]xFPGG that is directly upstream of the Nudix box (Figure 1.3), which is in contrast to the conserved NADH-motif located downstream of the Nudix box in previously discussed enzymes (Gasmi and McLennan 2001). The CoAse enzymes that have been studied in eukaryotes are mostly localized by a signal peptide to the peroxisome or the mitochondrial membrane (McLennan 2006). The peroxisome environment in particular can oxidize CoA to a dimer via disulfide bond, and therefore CoAses localized here can show a preference to oxidized CoA such as with Pcd1p of *S. cerevisiae* (Cartwright et al. 2000, McLennan 2006). The yeast CoAse was the first to be

studied, and its higher activity towards oxidized CoA was considered plausible based on the environment in which the enzyme was found. Later, the bacterial ortholog in *D. radiodurans* was isolated and characterized with the same preference, which again seemed plausible based on the extreme-living and resilience for which *D. radiodurans* is known (Kang et al. 2003). Such activity towards oxidized CoA fit the model of Nudix enzymes being involved in housecleaning, for CoA dimers can result from oxidative stress, and therefore lack the free thiol required for its acylation activity. With this hypothesis, a mechanism of CoA recycling was finally realized.

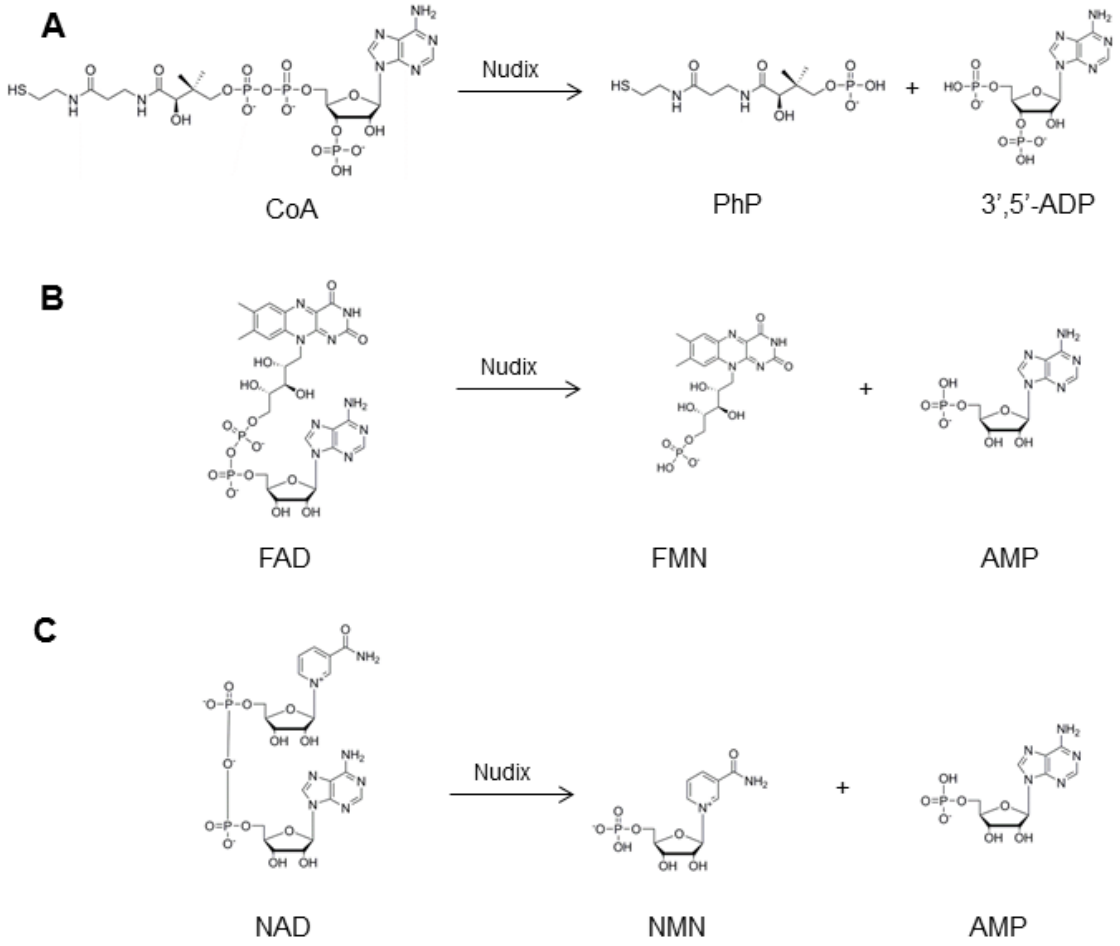


Figure 1.4 Nudix Hydrolysis of Coenzymes.

A, Hydrolysis of CoA to PhP and 3'5'-ADP known to be carried out by all specified Nudix CoAses with the exception of Human NUDT7 and E. coli NudL; B, Hydrolysis of FAD currently not carried out by any known Nudix enzyme; C, Hydrolysis of NAD currently carried out by E. coli NudC.

The discovery of Nudix CoAses in other organisms, however, provided evidence of other roles. The Nudix CoAse NDX8 in *C. elegans* was found to also be targeted towards the peroxisome, but did not show the same affinity for oxidized CoA as the previously characterized CoAse from *S. cerevisiae* (AbdelRaheim and McLennan 2002). While this enzyme was capable of

hydrolyzing the CoA dimer, the catalytic efficiency for CoA was higher. Studies on the Nudt7α enzyme from *M. musculus* that is likewise targeted to the peroxisome showed similar results that, while the affinity for oxidized CoA was slightly higher than CoA itself, the catalytic efficiency for CoA was almost two-fold greater for the reduced monomer than for the dimer (Gasmi and McLennan 2001). Further studies of Nudt7α showed that its catalytic efficiency increased with the size of CoA thioesters tested, and peaked at the C_{12} length of lauroyl-CoA with over 40-fold higher k_{cat}/K_m and 7-fold higher affinity than observed with CoA (Reilly et al. 2008). While still capable of cleaving oxidized CoA, these CoAses seem capable of so much more. The higher activity displayed against thioesters implicates the role of such enzymes in metabolic regulation, such as the switching between fatty acid metabolism and glycolysis, where CoA used in β-oxidation may be immediately required for use in respiration. The hydrolysis for lauroyl-CoA by Nudt7α, for example, would immediately free 3’5’-ADP as a direct substrate for CoA synthesis by PPAT and allow the cell to repurpose CoA into another pathway. In contrast, lauroyl-CoA (C_{12}) would require 20 enzyme-catalyzed reactions to convert completely to acetyl-CoA (C_2) via β-oxidation. Indeed, each subsequent two carbon removal from lauroyl-CoA would yield an acetyl-CoA molecule, but this would still require 4 enzymatic steps and the three coenzymes of CoA, FAD, and NAD to accomplish.

The putative Nudix CoAses offer the opportunity to further hone the model of CoAse function throughout evolution. The Conserved Domain Database shows over 140 occurrences of the Nudix motif with the CoA binding motif

throughout Bacteria, Eukarya, and Archaea (Marchler-Bauer et al. 2015). While there have been Nudix CoAses characterized in bacteria, yeast, insects, plants, and mammals, there are still a great number that remain putative (Kang et al. 2003, Cartwright et al. 2000, AbdelRaheim and McLennan 2002, Kupke et al. 2009, Gasmi and McLennan 2001, Reilly et al. 2008). The human proteins Nudt7 and Nudt8 are examples of such putative enzymes, for they show homology to the murine Nudt7α, but lack the peroxisome-targeting signal sequence (Gasmi and McLennan 2001). Therefore while it seems plausible that they both might maintain CoAse activity based on their sequences, the subcellular localization differs from the eukaryotic models presented by such organisms as *M. musculus* or *C. elegans*. Another example of a putative CoAse is YeaB (NudL) in *E. coli*. This protein contains both Nudix and CoA motifs similar to characterized enzymes in *D. radiodurans, M. musculus,* and *C. elegans* (Gasmi and McLennan 2001, McLennan 2006). Interestingly, despite the widespread availability of *E. coli* in laboratory settings and a transcript abundance of approximately one-sixth that of the indispensable dNTP-targeting NudB, this protein has yet to be characterized (McLennan 2006). Alexander McLennan and his group have done extensive research on the Nudix superfamily, and have referenced multiple times the possibility that this enzyme in *E. coli* may also target CoA, but to date there are no such characterizations of NudL. To date, the only group that has studied NudL has done so with focus on its promiscuous activity, wherein the enzyme was found to support the synthetic pyridoxal 5’ phosphate pathway with low activity (Kim et al. 2010). The authors recognize the activity is low, indicating this

is at best a secondary pathway, and mention that the predicted activity is towards CoA, however there is no such evaluation of NudL's substrate range outside of their focus on the pyridoxal 5' phosphate synthetic pathway. The only other *E. coli* Nudix enzyme that displays activity towards CoA is the aforementioned NudC. The study revealing this activity, however, was focused elsewhere, and only showed this NAD-targeting Nudix hydrolase to be qualitatively active against NAD, NADH and CoA without any investigation towards substrate range or kinetics, therefore leaving open the question of whether NudC functions as the *E. coli* CoAse or if the activity is another substrate promiscuity that seems characteristic of the Nudix enzymes (Bird et al. 2016, McLennan 2013). One reason this recent study failed to look into the possibility of NudC as a CoAse may be due to the fact some substrate range of NudC orthologs had already been probed, albeit the activity specifically towards CoA was not evaluated (Xu et al. 2000). If NudC is the principal *E. coli* CoAse, then it carries out this activity without the CoA motif (UPF0035) conserved in all other characterized Nudix CoAses throughout evolution (Figure 1.3). If there are multiple CoA-degrading Nudix enzymes in *E. coli,* the major enzyme would reasonably contain this CoA motif as is found in NudL. Nevertheless, despite the established characterization of a bacterial CoAse from *D. radiodurans* and the sequence data suggesting CoA hydrolysis activity by *E. coli* NudL, we still today are left without an established CoA-degrading enzyme in the otherwise well-characterized CoA pathway of *E. coli*.

The Nudix hydrolase superfamily has a logical place in evolution due to its antimutator functions. The importance of organismal self-preservation certainly warrants such a variety of enzymes that target and degrade potentially harmful nucleotide derivatives, and this is noted in early studies of the first characterized Nudix protein MutT (NudA) wherein knockout mutants displayed mutation rates up to 4-log higher than the wildtype (Bessman et al. 1996). This antimutator function, however, doesn't seem to fit when describing the coenzyme-targeting Nudix enzymes. The targeted coenzymes mentioned above contain no known mutation potential. In terms of NADH, the hydrolysis to produce NMNH is one potential function, but a downstream use has yet to be discovered for such a molecule (Frick and Bessman 1995). The preference for hydrolysis of NADH over NAD^+ can surely regulate the balance between the two, but for what reason would the cell destroy a molecule that has evolved to be cycled between the two states by such a variety of proteins? Destruction of NADH seems energetically wasteful, as the molecule could release its electrons for ATP production via electron transport in respiratory pathways. Indeed, the regeneration of NAD^+ is vital to fermentative metabolism, and the oxidation of NADH usually results in the production of some organic compound that may be excreted. However, cutting NADH to produce NMNH does not regenerate NAD^+ directly, primarily because there are no known biochemical reactions involving NMNH, especially those that would oxidize it to the NAD precursor NMN (nicotinamide mononucleotide, or nicotinamide D-ribonucleotide). The evolutionary reasoning behind the regulation of the NAD/NADH balance by hydrolysis is therefore unclear.

The function of CoA pyrophosphohydrolases similarly deviates from the Nudix superfamily. CoA has no known mutation potential similar to NADH. There is certainly a benefit to hydrolysis of dimerized CoA by enzymes such as Nudt7α, as it is biologically useless until it is reduced back to the free thiol-containing CoA monomer. Such Nudix activity may be especially important in organisms such as *S. aureus* and multiple *Bacillus* species (*B. cereus, B. subtilis, B. anthracis, B. thuringiensis*) that do not possess a glutathione redox system and use CoA in its place (Leonardi et al. 2005), however these organisms contain a CoA-disulfide reductase to regulate such oxidation. Furthermore, there are no known Nudix CoAses found in the Firmicutes phylum according to the CDD aside from 4 in the Clostridia family and 3 in Bacillaceae, and none of these include the previously mentioned CoA disulfide-reductase containing bacteria (Marchler-Bauer et al. 2015). In addition, *S. cerevisiae* that contains the oxidized CoA-preferring Nudix Pcdp1 enzyme does not contain a specific CoA-disulfide reductase enzyme. Therefore it seems these two enzymes are mutually exclusive. The role of Nudix CoAse as a recycler of oxidized CoA would be understandably reasonable in organisms that lack the specific CoA disulfide-reductase system, for organisms that contain glutathione do typically possess CoA-glutathione reductase enzymes that serve a similar purpose. However, the fact that many of the studied CoAses do not share the same preference for oxidized CoA as Pcdp1 from *S. cerevisiae* complicates this functional model. The idea that Nudix CoAses prefer CoA and even CoA thioesters above oxidized CoA while still capable of hydrolyzing them all implies these proteins function as multipurpose enzymes. As stated above,

the recycling of CoA by Nudix enzymes produces in a single step a direct substrate for CoA synthesis in 3’5’ADP, and such activity could have a place in the regulation of the energy state of the cell, especially when considering that excess CoA has been shown incapable of crossing the bacterial cell membrane unlike its precursor PhP (Jackowski and Rock 1984).

It’s clear that NAD- and CoA-targeting Nudix enzymes do not fit the role of antimutators like many of the other Nudix enzymes. These coenzyme-targeting Nudix enzymes still bear the Nudix motif and catalytic action, but they seem to be capable of much more than substrate turnover, simply due to the widespread impact of the substrates themselves. While CoAse activity is something never before demonstrated in model bacterial organisms such as *E. coli*, NAD degradation has been known since the 1970s by an unrelated enzyme (Grushoff et al. 1975). Now that there are two known coenzyme-targeting pyrophosphohydrolases from the same superfamily in organisms across multiple phyla, there is a question of what, if any, is their shared activity? The other Nudix hydrolases all seem to function in sub-classes of shared activity, but is this same type of association relevant to the coenzyme-targeting Nudix proteins? This question would remain wide open until almost 10 years after the Nudix CoAse was discovered.

1.4 Coenzyme-RNA Conjugates

1.4.1 In Vitro Conjugates

Conserved processes throughout evolution generally maintain the core mechanism, but gain specificity and regulation either with additional participating

molecules or additional protein components or subunits. One example of this idea is the RNA Polymerase (RNAP) and its catalysis of transcription. The RNAP maintains the same catalytic mechanism throughout all known life, but from the progression of bacteria to archaea and eukaryotes, it gains both additional subunits and binding partners to facilitate an initiation of transcription that is more complicated with evolution. Viruses, on the other hand, offer a simpler version of the process with a single RNAP subunit capable of carrying out transcription alone. This simple and efficient version of transcription makes viral RNAPs great candidates for in vitro transcription (Milligan et al. 1987). The in vitro transcription system can be utilized to produce RNA transcripts of any desired sequence downstream of a specific promoter. The T7 bacteriophage RNA polymerase became of particular interest in the early 2000s, when investigators discovered the advantage of the previously overlooked T7 φ2.5 promoter to allow transcription initiation with an adenosine nucleotide rather than a guanosine that was characteristic of the more commonly used T7 φ6.5 promoter (Huang et al. 2000). The impact of such a mechanism was rooted, for one, in the superior homogeneity of transcripts derived from the φ2.5 promoter that lacked the variable number of nucleotides at the 5' end (Coleman et al. 2004). Second, the installation of adenosine as the initiating nucleotide of transcription allowed the efficient incorporation of an endless potential of adenosine derivatives as initiators of transcription so long as Watson-Crick base pairing was not disrupted (Huang et al. 2003, Li et al. 2005, Huang and Shi 2010, Huang and Shi 2012) (Figure 1.5). In addition, biologically relevant adenosine derivatives such as CoA,

NAD, and FAD were capable of transcription initiation under the φ2.5 promoter, and this aspect brought forth in vitro coenzyme-RNA conjugates (Huang 2003a). The installation of nucleotide derivatives such coenzymes into RNA had been noted previously using the E. coli RNAP, however these observations were limited to incorporation into abortive transcripts of 2 nucleotides (Malygin and Shemyakin 1979). The discovery by the Huang lab demonstrated a far superior system wherein transcripts of any length and any defined sequence can be initiated with any adenosine derivative in high yield.

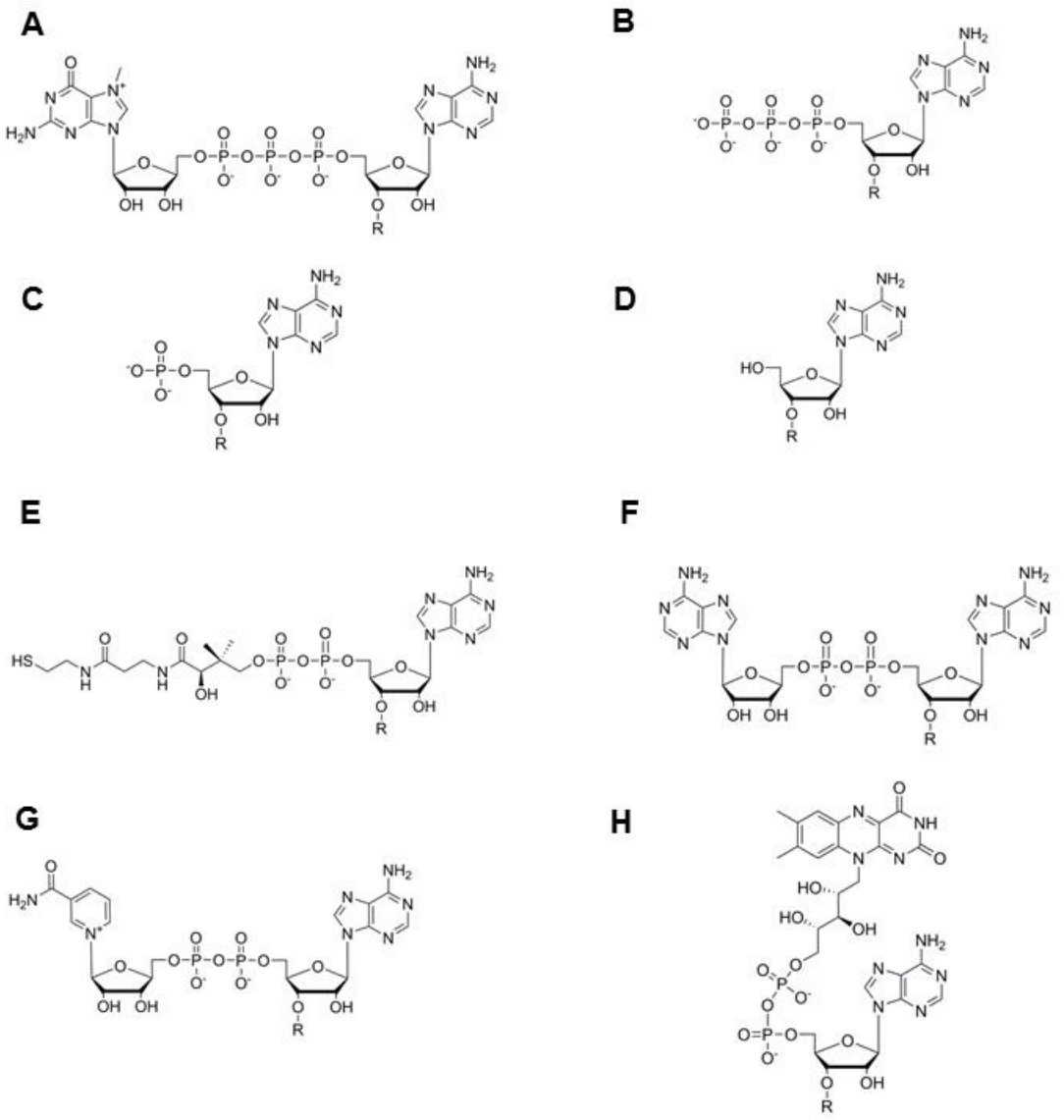


Figure 1.5 RNA Modifications.

Currently known RNA modifications that can be made in vitro and speculated in vivo. A, 7-methyl guanosine mRNA cap from eukaryotes; B, 5'-triphosphate cap of bacterial mRNA; C, 5'-monophosphate RNA as a degradation product of bacterial mRNA; D, 5'-hydroxyl RNA generated in vitro; E, CoA-RNA generated in bacteria; F, AppA-RNA, generated in vitro and speculated possible in bacteria; G, NAD-RNA generated in bacteria; H, FAD-RNA generated in vitro and speculated in bacteria.

The *in vitro* transcribed RNA conjugates have had many uses since their discovery. For example, the conjugation of fluorescent dyes to adenosine allow for fluorescent labeling of RNA is useful in localization studies in cell culture, or

non-radioactive hybridization techniques reminiscent of the Northern blot (Huang et al. 2003, Li et al. 2005). As RNA therapy has become an avenue for cancer treatment, RNA modifications that enhance stability during delivery to cell culture are vital (York et al. 2009, Huang et al. 2008). Such delivery and treatment enhancements can be achieved by the addition of such fluorescent derivatives or thiol derivatives, for immune responses to targeted siRNA knockdown therapy can be avoided by replacing the 5'-triphospate of RNA produced with *in vitro* transcription, (Hornung et al. 2006). In addition to altering the biological behavior of RNA molecules with modified 5' termini, the derivatization of adenosine with a free thiol introduces the potential for unique chemistries otherwise unavailable with canonical RNA. For example, 5'-thiol labeled RNA can be covalently attached to the surface of gold nanoparticles that are commonly used for RNA delivery in therapeutic applications (Huang and Shi 2010, Boisselier and Astruc 2009). The thiol modification can also be used for immobilization on solid surfaces by the formation of disulfide, thioether, or thioester bonds. Such immobilization is vital in the process of Systematic Evolution of Ligands by Exponential Enrichment (SELEX), wherein small populations of sequences that contain unique folding capabilities to facilitate high affinity binding (aptamers) or catalytic activity (ribozymes) can be isolated from large pools of randomly sequenced oligonucleotides (Tuerk and Gold 1990).

Coenzyme A can serve as an alternate source of a free thiol group on RNA if used as a transcription initiator as opposed to thiol-modified adenosine (Figure 1.5E). In addition to the chemistry advantages of the thiol group, CoA

incorporation can increase the biological relevance of SELEX results. In other words, thiol-labeled adenosine is useful in immobilization for the purposes of selection, but the activity of the resulting ribozyme is due to the sequence elements and not the thiol itself. On the other hand, CoA-conjugated RNA can be used as a way to immobilize the RNA, or as an active part of the ribozyme activity. In this regard, CoA-RNA has been used to demonstrate ribozyme activity directly involving the coenzyme (Coleman and Huang 2002). The transcription of CoA-conjugated RNA and subsequent discovery of RNA-catalyzed acylation activity is evidence that strongly supports the RNA world theory and the evolution of coenzymes (White 1976). The acylation of CoA, as described above, is a major component of cellular fatty acid metabolism and one of the major functions of contemporary CoA. While such activity catalyzed by RNA had been hypothesized for over 40 years, the discovery of such a ribozyme was not possible without the advances of RNA-conjugate synthesis.

In vitro conjugates of RNA have proven useful in the recent 20 years. However their uses for imaging, detection, and delivery are due to unnatural modifications. The advantages offered by such modifications supports the potential applications of RNA, but well-planned unnatural additions could certainly increase the scope of use for any molecule. RNA modifications that allow for immobilization and result in discovery of new ribozymes could easily be equated with small molecule modifications used for immobilization and discovery of aptamers by the same process. The discoveries by SELEX alone made possible by such RNA conjugates are indeed impressive and provide data for the

ongoing debate of the origin of life. As such, these modifications exist as tools, with little to no biological significance themselves. On the contrary, coenzyme-RNA conjugates surpass the point of being just tools to achieve a goal. Not only is this modification constructed by natural means as are the others, but the modification itself is found in all organisms in high concentration. The discovery of a CoA-linked ribozyme capable of acylation supports a 40 year old theory of coenzyme origin as a functional remnant of a ribozyme, and demonstrates a biological significance to coenzyme-RNA conjugates (White 1976, Coleman and Huang 2002). While there are no current examples of NAD- or FAD-linked ribozymes, a redox-catalyzing ribozyme was discovered over a decade ago that was shown to interact with NAD^+ for its alcohol dehydrogenase activity (Tsukiji et al. 2003). While this ribozyme is closer related to modern redox enzymes that interact with NAD^+, undiscovered NAD-linked ribozymes are likely capable of the same activity just as CoA-linked ribozymes are capable of acylation. Until the discovery of an NAD-linked ribozyme is made, the biological application of this *in vitro* coenzyme-RNA conjugate is still questionable. If and when such a ribozyme is discovered, the biological relevance of this conjugate, similar to the acylating CoA-ribozyme, will be an evolutionary one. In order to extend beyond just an evolutionary impact, evidence of the natural existence of coenzyme-conjugated RNAs in contemporary organisms is needed. Such a discovery would certainly strengthen the argument for the natural existence of coenzyme-linked ribozymes, but it also would redefine our understanding of the role of RNA in contemporary organisms. Fortunately, less than a decade after the discovery of CoA-

ribozymes, evidence was produced that transitioned the coenzyme-RNA relationship from a strictly evolutionary setting to a contemporary one.

1.4.2 In Vivo Conjugates

The concept of coenzyme-linked RNA has been around for many years. It was hypothesized by White in 1976 with his proposed origin of coenzymes, and demonstrated by on a small scale in *E. coli* three years later (White 1976, Malygin and Shemyakin 1979). The observation by Malygin that coenzymes could be incorporated as initiators into abortive transcripts was significant, but the impact was wanting. Does incorporation into abortive transcripts signify some biological significance? After all, abortive transcripts result when the polymerase is incapable of transitioning into the elongation phase of transcription. Is the result simply from aberrant polymerase activity that is corrected by transcription abortion? Is this any different than the simple adenylation of a coenzyme? Is this a conserved mechanism with contemporary value or is this another vestige of an obsolete era? Decades later when coenzyme incorporation into longer RNA was accomplished with the T7 φ2.5 promoter, some of these questions were wiped away, but the natural occurrence in contemporary organisms was still unknown (Huang 2003a). Despite the impact of the T7 promoter system allowing such incorporation into RNA, the results were still *in vitro* synthesized molecules used for an *in vitro* technique to isolate non-naturally occurring ribozymes reminiscent of an era long passed. However concrete the evidence was in terms of sequence knowledge and kinetic data, the entire concept itself was still somewhat abstract without some link to a contemporary organism.

A set of discoveries in 2009 opened the door to areas of study and pathways never before known to exist. In two studies from the same lab, RNA was isolated from *E. coli* and *S. venezuelae* that was determined to be covalently linked to the coenzymes NAD and CoA (Kowtoniuk et al. 2009, Chen et al. 2009) (Figure 1.5E and 1.5G). Fractionation of total RNA from these bacteria indicated the coenzymes were incorporated into smaller sized RNA 200 nucleotides or less in length, and the identities of the conjugates were determined by LC/MS analysis. In addition to CoA, RNA was found linked to CoA thioesters, specifically succinyl-CoA, methylmalonyl-CoA, and acetyl-CoA (Kowtoniuk et al. 2009). These discoveries together serve as the first demonstration of *in vivo* coenzyme-RNA conjugates. Such a discovery strengthens the arguments for the existence of coenzyme-containing ribozymes that previously were unanchored to any organism existing after the LUCA. In addition, these reports generate questions about the function, the abundance, the degree of conservation, the diversity of different coenzymes used, the synthesis mechanisms, the biodegradation mechanisms, and the dispersion of coenzyme-RNA across other phyla. RNA modifications such as these have never been observed or even hypothesized outside of an evolutionary context that predates the protein world, and therefore these questions may constitute a new field of study in a similar manner to the discovery of microRNA.

The diversity of the coenzyme-RNA conjugates is an interesting topic that may be related to function and dispersion across different phyla. The discovery of CoA thioester-linked RNA is particularly interesting because of the variety of

thioesters known to exist. For example, the methylmalonyl-CoA RNA was only discovered in *S. venezuelae*, purportedly due to the lack of methylmalonyl-CoA synthesis in *E. coli* (Kowtoniuk et al. 2009). This observation implies that not only is the process itself conserved, but the synthesis of RNA modification may be modified to suit individual species based on their metabolic capabilities. In addition, the discovery of such RNA in the distinct phyla of Actinobacteria and γ-γ-Proteobacteria implies a widespread dispersion across the entire domain, for these phyla are two of the furthest diverged in all of Bacteria according to recent phylogenomic analysis (Wu et al. 2009). The possibility of distinct CoA-RNA pathways that depend on the energy state of the cell in addition to the thioester synthesis capabilities of an organism is intriguing, and could represent an entire area of metabolism and regulation previously unknown. Furthermore the detection of distinct CoA-thioester modifications in this case may serve as a biomarker. On the other hand, such CoA thioesters may be incorporated into RNA by a single conserved enzyme with a wide substrate range, and the different CoA thioesters identified may represent the cellular CoA/thioester ratio reflective of growth conditions and energy state (Leonardi et al. 2005). If these different CoA thioesters are installed as a reflection of the metabolic state, then what purpose do they serve? Is there some selective advantage of one thioester over another? Some basic knowledge of coenzyme-RNA function is required to give direction to the reasoning behind specific coenzymes or even thioester modifications.

The function of coenzyme-RNA conjugates has a high potential to be related to the sequence and function of the RNAs themselves. It's hard to imagine CoA and NAD incorporation to be by way of aberrant transcription initiation or coincidental post-transcriptional conjugation. Given the significance of both coenzymes in metabolism, it's more reasonable that the coenzymes play distinct roles tied to either the function of the RNA or the significance of that RNA. Whether installed via transcription or post-transcriptionally, both options would require a novel system of catalysis and regulation to complete. The investigators unsuccessfully attempted to replicate the installation of the coenzymes using *E. coli* RNA polymerase, therefore leading them to believe this modification was a post-transcriptional modification of the RNA (Kowtoniuk et al. 2009). If this is true, then what mechanism is in place to distinguish one sequence from another to divert to post-transcriptional modification? Transcription and translation are understood as coupled processes in that occur concurrently, but does the size of this RNA exclude it from being assumed as an mRNA? Even if only half of the approximately 200 nucleotide RNA was used as an open reading frame, it would hypothetically yield a peptide with 33 amino acids that is roughly 7 kD. Therefore the size should not preclude the possibility of it being mRNA.

The sequences associated with the coenzyme-RNA conjugates are of great importance in determining the functions of these molecules, however they were unfortunately unobtainable in the original publication. The method used to isolate and identify the conjugates involved nuclease treatment to produce single

nucleotides before LC/MS analysis, thereby destroying any trace of the associated sequences (Chen et al. 2009, Kowtoniuk et al. 2009). A method was eventually developed to capture intact NAD-RNA, and the sequence information obtained indicated a significant portion of the RNA were regulatory sRNA (Cahova et al. 2015). Interestingly, the most abundant NAD-RNA identified were RNAI and CopA, which have been characterized as antisense regulators of ColE1 and R1 plasmid regulation, respectively (Wagner and Simons 1994). In fact, 12 of the 15 known sRNA that were identified act as antisense regulators. Aside from RNAI and CopA, the other identified NAD-sRNAs have previously been characterized as regulatory of stress responses, such as acid stress and low magnesium (Cahova et al. 2015). In addition to sRNA, sequencing also revealed NAD linked to 5'-termini of mRNA encoding metabolic enzymes such as tagatose-1,6-bisphosphate aldolase (gatY), glycerol-3-phosphate dehydrogenase (glpA and glpD), methylglyoxal synthase (mgsA), and threonyl-tRNA synthetase (thrS) among others (Cahova et al. 2015). The mRNA of multiple transcription and translation regulators were also identified, but similar to previous examples, the sequences only mapped to the 5'-termini and not the full-length transcript. For some of the identified genes, the existence of the 5'-termini alone can be easier understood, such as those containing leader peptides to hisL and ilvL that are known to attenuate the operons for amino acid synthesis. The reasoning for this truncation of the identified mRNA is unknown and can only be speculated. Perhaps the truncation is enough to allow feedback inhibition of these genes similar to the translation of the inhibitory leader peptides for amino acid

synthesis. However it is unknown whether the NAD-modification inhibits or enhances the translation of the associated mRNA.

Based on the abundance of sRNA identified that regulate stress responses, the investigators hypothesized the NAD-conjugation to act as a cap of sorts that offers stabilization of these RNAs in the cell. Bacterial RNA is largely known to contain a 5'-triphosphate group that offers protection from the degradation pathway catalyzed by both endonuclease and exonuclease enzymes. In *E. coli*, this pathway is initiated by RNaseE that requires interaction with a 5'-monophospate to initiate internal cleavage (Bandyra et al. 2012). Therefore the presence of a 5'-triphosphate group can prevent RNaseE activity from initiating RNA degradation. This triphosphate "cap" can be removed by the activity of RNA 5' Pyrophosphohydrolase (RppH) that generates 5'-monophosphate-RNA and PPi, therefore enabling RNaseE-initiated degradation to occur (Bandyra and Luisi 2013). The addition of a coenzyme to the 5'end of RNA therefore makes resistant to RppH, adding stability to the RNA (Cahova et al. 2015). Considering that an RNA cap other than a triphosphate was never known before in bacteria, a decapping bacterial enzyme was also unknown. Cahova hypothesized that the activity of NudC, an NADH-hydrolyzing Nudix enzyme, would serve to "decap" NAD-RNA and push it to the previously characterized degradation pathways. It is not uncommon for Nudix enzymes to act as RNA decappers, for members of the Nudix family from various organisms have the ability to hydrolyze the eukaryotic 7-methyl guanosine (7mG) mRNA cap such as human NUDT20, yeast NUDT16, and viral D10 among others

(McLennan 2006, Song et al. 2013, Parrish et al. 2007). The decapping is notably different between NudC and others such as NUDT20 considering that the structures of the eukaryotic 7mG cap and the bacterial NAD-RNA cap are structurally dissimilar, but they all function to initiate RNA decay pathways.

The degradation of coenzyme-RNA in bacteria is potentially different from mRNA decapping in eukaryotes due to the existence of multiple structurally distinct bacterial coenzyme-RNA conjugates. Therefore the demonstration of NudC activity in bacteria would be specific for NAD-RNA, but not CoA-RNA (Cahova et al. 2015). While the first characterization of NudC in the 1990s included activity across a range of substrates, activity against other coenzymes was not addressed (Frick and Bessman 1995, Xu et al. 2000). However, investigators recently demonstrated that NudC appears capable of releasing both NAD-RNA and CoA-RNA *in vitro* (Bird et al. 2016). As discussed earlier, while NudC and its orthologs were capable of hydrolyzing substrates such as dNTPs, this activity was 2 orders of magnitude or more lower than that observed for NADH (Xu et al. 2000). The difference in activity of NudC against NADH and NAD^+ was shown to be around 5-fold, and therefore the hydrolysis of a structurally distinct coenzyme such as CoA would be expected to be comparably low (Xu et al. 2000). Based on the differences in activity expected between CoA and NAD hydrolysis by NudC, the lack of a conserved CoA binding motif in NudC, the lack of characterization on CoA hydrolysis by NudC in recent investigations, and the lack of evidence that NudC is capable of hydrolyzing CoA thioesters, it is unclear whether the release of CoA-RNA would be catalyzed *in*

vivo by this enzyme at an acceptable rate to be considered responsible for CoA-RNA “decapping”. On the other hand, members of the Nudix family have been characterized in many organisms that specifically hydrolyze CoA and its thioesters. Based on the sequence predictions of *E. coli* NudL and the activity of its orthologs, it’s highly likely that this is the enzyme responsible for turnover of CoA-RNA. Unfortunately, however, this enzyme has not yet been and characterized for its activity against CoA, and therefore remains a only putative CoAse (McLennan 2006).

The RNA sequences associated with CoA remain unknown. In fact, little about CoA-RNA is known aside from its existence. With such little information, the subject of its function, synthesis, degradation, and modification all remain topics of speculation. While the function of such RNA is hypothesized to be related to NAD-RNA, it seems unlikely that two distinct structures would exist for the same purpose. Indeed, both coenzymes would offer protection against RppH-initiated degradation of the associated RNA, but the functions of the two likely extend beyond just RNA protection. NAD and CoA are known to associate with a variety of different proteins and enzymes, and their conjugation to RNA would increase a transcript’s protein association potential. Such an ability to associate with a variety of proteins would offer another means of regulation, wherein proteins can aid in RNA transport or alter translation rates. In addition, the RNA sequence itself could serve as a docking site to facilitate and stabilize protein-protein interactions. One caveat to the comparison of NAD-RNA and CoA-RNA function is the variations of known CoA-RNA. NAD-RNA has only been

discovered in the oxidized form of the coenzyme, but CoA itself in addition to three separate thioesters have been isolated and identified. These four different forms of CoA represent different functional states used throughout metabolism whose concentrations vary based on the current cellular environment. The diversity of CoA found in RNA-conjugates could indicate not only a function distinct from that of NAD, but a distinct use for each CoA derivative conjugated to RNA. Such a hypothesis depends of course on the mechanism of CoA installment on the RNA, and whether that mechanism is universal for CoA, CoA thioesters, and NAD. Investigators have demonstrated incorporation of CoA and NAD into RNA by both *E. coli* and T7 RNA polymerases, but no one has yet demonstrated such a process with a CoA thioester (Huang 2003a, Bird et al. 2016). While it is possible CoA thioesters can be installed similarly, it is also possible that CoA-RNA itself is capable of acylation to produce these thioester CoA-RNA conjugates. The mechanism of CoA installation is further complicated by the issue of depCoA. The demonstration of CoA-RNA transcription *in vitro* requires depCoA. As the substrate of PPAT and the precursor to mature CoA, depCoA is not known to be widely available in *E. coli* and is also not well studied in terms of intracellular concentrations. In eukaryotes, depCoA is expected to be even less abundant due to its production and subsequent consumption by the dual-functioning enzyme CoA Synthase (Leonardi et al. 2005). Given that 4’phosphopantetheine is the major intracellular intermediate of CoA synthesis in *E. coli*, it’s likely that CoA-RNA may be synthesized with PhP and ATP-RNA by a similar activity to that of PPAT (Jackowski and Rock 1981). Recent studies

identified the most abundant NAD-RNA shared promoter elements that initiated transcription with adenosine, and were able to replicate installment of depCoA, NAD^+, and NADH *in vitro* with *E. coli* RNA polymerase and σ^{70} (Bird et al. 2016). The investigators were also able to transform *E. coli* with a plasmid encoding a transcript under the same promoter to demonstrate the decapping capabilities of NudC on isolated RNA, concluding that these structures were incorporated via transcription initiation. The promoter used would indeed install ATP as the initiating molecule, and while the investigators argued the molar ratios of coenzyme to NTP needed to be tuned to facilitate coenzyme initiation *in vitro*, there is no such control *in vivo* to produce similar results. Such a disparity may support the hypothesis that such transcripts were initiated with ATP and were then subsequently modified with PhP by PPAT-like activity to form CoA-RNA. The same basic reaction can be hypothesized for other coenzyme-RNAs as well based on a previously discovered ribozyme capable of catalyzing the reaction of R-p + ppp-A → R-ppA + PPi, wherein the reaction of ATP-initiated RNA and the coenzyme precursor NMN could be catalyzed to form NAD-RNA as well (Huang et al. 2000). Such a mechanism may explain why NADH-RNA was never initially discovered and has still not been confirmed by subsequent investigators (Chen et al. 2009).

In addition to the unknown synthesis and functions of CoA-RNA, it is unclear whether such RNA exists anywhere else. It's clear that Nudix enzymes capable of similar coenzyme degradation exist in many organisms, and it's been demonstrated that simple RNA polymerases can install such molecules into

transcripts, but does this RNA modification extend beyond the domain of Bacteria? If such RNA modifications do exist in other domains, do they serve the same purpose as those proposed in Bacteria? Furthermore, are there other coenzyme-RNA conjugates yet to be discovered? The most likely candidate so far overlooked as a coenzyme-RNA conjugate is FAD. This coenzyme is structurally similar to NAD but with an isoalloxazine group instead of nicotinamide, and is known to be capable of initiation of transcription using T7 RNA polymerase (Huang 2003a). FAD can also be synthesized by the previously mentioned ribozyme reaction of R-p + pppA → R-ppA + PPi. While there is no characterized Nudix enzyme whose primary substrate is FAD, the human NUDT12 and *E. coli* NudE have both been demonstrated to hydrolyze this coenzyme (McLennan 2006, O'Handley et al. 1998). Therefore if the synthesis and degradation capabilities are in place, there is a possibility that FAD-RNA exists and has yet to be discovered.

The length of time between the discovery of coenzyme-RNA in 2009 and the identification of NAD-RNA sequences in 2015 is an indication of the technical barriers associated with this area. The identification alone of the coenzymes required destruction of the RNA itself, leaving other investigators marveling at the potential of coenzyme-RNA conjugates while presumably stifled in their attempts to learn more. The eventual capture of NAD-RNA required a specific chemistry wherein the nicotinamide group was enzymatically replaced with an alkyne via transglycosylation that could undergo copper-catalyzed azide-alkyne cycloaddition to a biotin derivative, therefore allowing immobilization of the NAD-

RNA on streptavidin resin (Cahova et al. 2015). The drawback to such an efficient capture protocol is that the initial enzymatic transglycosylation requires the reactive N-glycosidic bond found in the oxidized form of NAD^+, and therefore excludes any other known or hypothesized coenzyme-RNA conjugate. The likewise capture protocol of CoA-RNA has yet to be discovered, leaving the sequences of such RNA a mystery and the questions of function unanswered.

The identification of CoA-RNA and the elucidation of its function and metabolism are undertakings associated with significantly unique problems. The capture of CoA-RNA will require a specific protocol to be developed just like what was required for NAD-RNA identification. The analysis of sequences will require bioinformatics studies to place the function of these sequences to hypothesize a function for CoA-RNA. The determination of metabolic components involved in synthesis and degradation may reveal previously unknown enzymes and pathways that themselves have never been characterized. Despite these potential roadblocks and other problems associated with the development of methods for each aspect of CoA-RNA function and metabolism as discussed earlier, the story behind CoA-RNA can still be obtained. Many fundamental questions may be given new perspectives from which we can further tune our hypotheses about the origin of life. Are there more functions for CoA to add to the myriad of responsibilities CoA has in the contemporary cell? Has life evolved divergent pathways for modifying bacteria with ancient biomolecules for selective advantages? Does this addition to the repertoire of RNA modification extend beyond the domain of Bacteria? Can these coenzyme-RNA molecules still

function similar to naked coenzymes in the cell, and therefore exist as evolutionary intermediates of RNA World ribozymes and the contemporary living fossils of coenzymes? Do these RNA sequences bear some catalytic activity related to the function of their coenzyme conjugates, therefore existing as contemporary ribozymes previously accepted as supplanted by protein enzymes in evolution? Some of these questions may never be known, but the idea of science in general is the addition of knowledge to the collective global database of information, so that we can better understand ourselves, our environment, our history, and our future. Jerry Joyce shared the idea in 1989 that due to the complexity of life's tumultuous history on earth, we can only base our hypotheses towards understanding the different of aspects life on what little evidence we can gather, and as new experimental evidence is produced over time, we cannot help but view such questions from different perspectives. The uncovering of CoA-RNA function and metabolism is one such feat that will certainly inspire new perspectives.

CHAPTER II - OBJECTIVES OF RESEARCH

The recent discoveries of bacterial coenzyme-RNA and the sequences associated with NAD-RNA have opened a new area of study that requires unique methods to address. The existence of such *in vivo* RNA modifications is reminiscent of molecules hypothesized in the RNA World theory, and therefore their contemporary functions likely represent a snapshot of the evolutionary transition to the contemporary protein world that, much like coenzymes themselves, continue to bestow a selective advantage to organisms employing them. While some work has been done to identify sequences associated with NAD-RNA and hypothesize its function and metabolism, no such work has been carried out on CoA-RNA. Due to the unique chemistry of which CoA is capable, the maturation of modern cloning techniques, and the power of Next Generation Sequencing, we have a unique opportunity to take advantage of these advances to develop a capturing protocol and establish the function and metabolism of CoA-RNA.

We hypothesize that CoA-RNA can be captured by thiol chemistry and specifically released in a pure form by recombinant Nudix enzymes, wherein the resulting RNA can be prepared for Next Generation Sequencing for identification. We further hypothesize that recombinant expression and characterization of *E. coli* NudL will establish its role as the CoA pyrophosphohydrolase responsible for CoA-RNA turnover. The overall goal of this research is to develop a method to capture CoA-RNA for the identification of its function, and shed light on its

synthesis and degradation pathways in *E. coli*. The specific objectives of this research are:

1. **Cloning of Nudix Hydrolases with potential CoA Hydrolase activity.** The specific release of CoA-RNA from a capture protocol while preserving the nucleotide sequence relies on the specific hydrolysis of CoA itself. The Nudix hydrolase superfamily contains several characterized CoA pyrophosphohydrolases that are capable of such activity from different organisms. Candidate enzymes will be cloned for recombinant expression in *E. coli* for use as the release step of CoA-RNA purification. Due to differences in protein solubility, affinity, and specific activity of these enzymes, multiple candidates from different organisms will be cloned and characterized for their ability to release CoA-RNA from solid-state resin.
2. **Establishment of *E. coli* NudL as a CoA pyrophosphohydrolase.** Characterized Nudix enzymes with CoA pyrophosphohydrolase activity from all organisms share two amino acid sequence elements: a Nudix box and a CoA binding motif. These two consensus sequences have allowed such CoAse enzymes to be predicted in a variety of organisms including *E. coli*. Given the recently established importance of CoA in *E. coli* as an endogenous RNA conjugate, and the propensity for *E. coli* to be used as a model organism due to its widespread availability and accommodating culturing conditions, it is peculiar that such an enzyme has been predicted for at least a decade but never characterized. The enzyme YeaB, recently renamed NudL, will be cloned and purified as a soluble recombinant

protein from *E. coli*, whereupon it can be characterized as a CoA pyrophosphohydrolase with various activities towards CoA derivatives such as oxidized CoA, CoA thioesters, and CoA-RNA.

3. **Assessment of the coenzyme substrate ranges of Recombinant Nudix enzymes.** Members of the Nudix hydrolase superfamily are known for maintaining a large range of substrates. As the catalytic activity is specific for hydrolysis of a phosphoanhydride bond, many molecules containing such a bond can be hydrolyzed by Nudix hydrolases at different rates. In addition to the CoA-specific sequence elements of NudL, other *E. coli* Nudix hydrolases with sequence elements specific for other coenzymes related to NAD have been shown to cleave CoA. If the NudC enzyme is capable of hydrolyzing CoA in a similar manner to NAD and NADH, for example, then it likely can also hydrolyze FAD and $FADH_2$. Therefore such Nudix enzymes will be cloned and expressed as recombinants so that their activity against other coenzymes may be uncovered, and to establish evidence to support the assertion of NudC as a general coenzyme hydrolase in *E. coli.*
4. **Development of CoA-RNA capture protocol for the identification of CoA-RNA sequences.** The capture of CoA-RNA from total *E. coli* RNA will require specific thiol chemistry and specific enzyme-catalyzed CoA hydrolysis. Such an undertaking has never before been established, and therefore a protocol must be developed that allows the purification and enrichment of such a rare RNA modification while also allowing its

manipulation to generate a Next Generation Sequencing (NGS) library with minimal enzymatic steps to maintain RNA quality. The constructed libraries can be subjected to bioinformatics analysis via NGS by referencing the established *E.coli* genome and annotations to allow the identification of the CoA-RNA sequences and their abundance. Taken together with the data obtained by Cahova et al. (2015) on NAD-RNA, functional overlaps and divergence of coenzyme-RNA can be established.

5. **Determination of CoA-RNA metabolism in *E. coli.*** The establishment of NudL as a specific CoA pyrophosphohydrolase in *E. coli* insinuates its participation in CoA-RNA hydrolysis *in vivo*. While previous investigators have attempted to show that the activity of another Nudix enzyme, NudC, is responsible for *in vivo* NAD-RNA and CoA-RNA hydrolysis, their conclusions are empty without a specific CoA-RNA capture protocol. By identifying suspected CoA Hydrolases in *E. coli,* the metabolism of CoA-RNA can be probed by processing total RNA with the CoA-RNA capture protocol from *E. coli* overexpressing NudC and NudL individually, and from *E. coli* knockout mutants deficient in these enzymes to determine the difference in yield from controls.

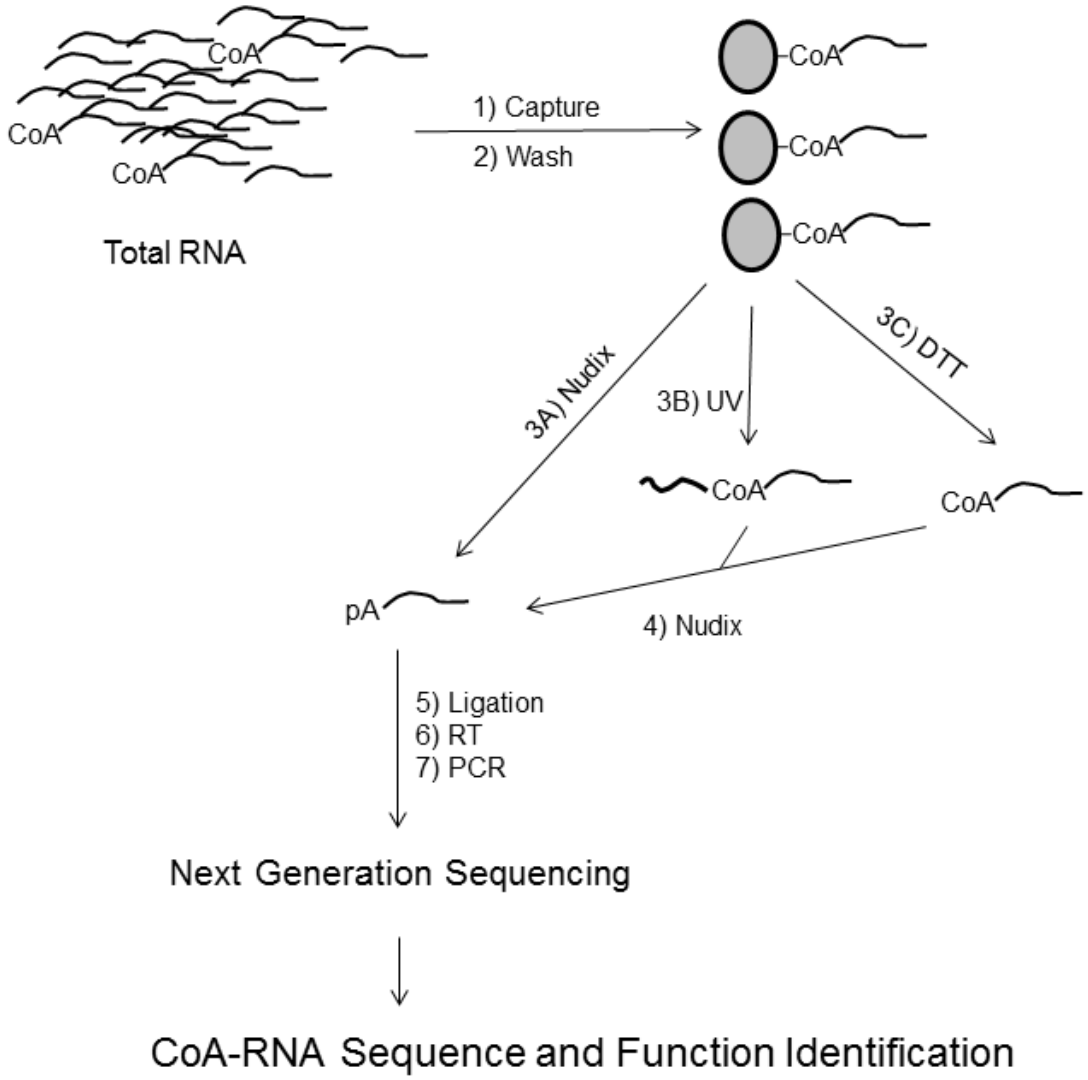


Figure 2.1 Project Overview.

The isolation of CoA-RNA requires the discriminatory activity of specific Nudix CoAses for isolation (Steps 3A and 4). Therefore the isolation and characterization of Nudix CoAses will be completed first, followed by the specific capture of CoA-RNA from total RNA and subsequent preparation of NGS libraries for the identification of CoA-RNA sequences and functions.

The following chapters describe the successful completion of these objectives. The cloning of Nudix CoA pyrophosphohydrolases, the establishment of NudL as a CoA pyrophosphohydrolase, and the assessment of coenzyme substrate ranges among Nudix enzymes are described in Chapter IV. The

development of a CoA-RNA capture protocol for the identification of CoA-RNA sequences is described in Chapter V.

CHAPTER III – MATERIALS AND METHODS

3.1 Instrumentation and Equipment

Name – Model (Company)

Accumet Basic pH meter – XL25 (Fisher Scientific)

Analytical Balance – HR 60 (A&D Co)

Autoclaves – LG 250 and 3031-S (Amsco Eagle)

Balance – Pioneer (Ohaus)

Centrifuge – Avanti J-26 XP (Beckman Coulter)

Centrifuge Rotors – JLA 16.250 and JA 25.50 (Beckman)

Electrophoresis Power Supply – EC 154 (E-C Apparatus Corporation)

Electrophoresis Power Supply – EC4000P (E-C Apparatus Corporation)

Electrophoresis Power Supply – PowerPac Basic (BioRad)

Freezer -20 °C (Frigidaire)

Freezer -80 °C (Forma Scientific)

Geiger Counter – Model 3 (Ludlum Measurements Inc.)

Gel Documentation System – ChemiDoc MP (Bio-Rad)

Gel Dryer – Model 853 Gel Dryer (Bio-Rad)

Heating Blocks – 12359030 (VWR)

HPLC Control - SCL-10A VP (Shimadzu)

HPLC Diode Array Detector – SPD-10A VP (Shimadzu)

High Pressure Liquid Chromatograph – LC-10AT VP (Shimadzu)

HPLC Degasser – DGU 14A (Shimadzu)

Horizontal Electrophoresis System – FB-SB-710 (Fisher)

Horizontal Electrophoresis System – Mini-sub Cell GT (BioRad)

Incubator 37 °C – 120 (Labline Instruments Inc.)

Liquid Scintillation Counter – 425 034 (Hidex)

Microcentrifuge – 5415 C (Eppendorf)

Microplate Reader – SpectraMax M3 (Molecular Devices)

Micron Centrifugal Filter Devices – M3, M10, M30, & M50 (Millipore)

Phosphorimager/Molecular Imager FX – 729 (BioRad)

Pipettes – P-1000, P-200 (Gilson)

Pipettes – P-10, P-2 (Fisherbrand)

Pipettes – P-200, P-20 (Thermo)

Refrigerator -20 °C & +4 °C (Kenmore)

Sonicator – XL-2000 (Misonix)

Spectrophotometer – 8453 (Agilent)

Stirrers (Nuova)

Thermocycler – Mastercycler (Eppendorf)

UV Lamp – Blak-Ray B100A (UVP)

Vertical Gel Apparatus – V16 (GibcoBRL)

Vertical Gel Apparatus – 237,504-7 (Sigma)

Vertical Gel Apparatus – Mini Protean 3 (BioRad)

Vortexer – G560 (VWR)

3.2 Chemicals, Reagents, and Kits

Name (Company)

Acylamide (40%, 19:1) (BioRad)

Acrylamide (30%, 37.5:1) (BioRad)

Adenosine triphosphate (Sigma)

Agar (DIFCO)

Agarose (Fisher)

Ammonium Persulfate (Sigma)

Ampicillin sodium salt (Sigma)

Ampliscribe T7 Flash Transcription Kit (Epicentre)

Amylose Resin (New England Biolabs)

BCA Protein Assay Kit (Pierce)

Beta-Mercaptoethanol (Sigma)

Beta-Nicotinamide Adenine Dinucleotide Hydrate (Sigma)

Boric Acid (Sigma)

Bovine Serum Albumin (BSA) (Fisher)

Brilliant Blue G (Sigma)

Bromophenol Blue (Sigma)

Chloramphenicol (Research Products International Corp.)

Coenzyme A Trilithium salt (Sigma)

Dextrose (Fisher)

DTT (Sigma)

EDTA Disodium Salt (Fisher)

Ethanol (Decon)

Ethidium Bromide (Sigma)

Expresso T7 Cloning & Expression System (Lucigen)

Expresso T7 SUMO Cloning & Expression System (Lucigen)

Flavin Adenine Dinucleotide Disodium Salt Hydrate (Sigma)

GelCode Blue Safe Protein Stain (Thermo)

Glacial Acetic Acid (Sigma)

Glycerin (Fisher)

Glycine (Research Products International Corp.)

Hydrochloric Acid (Fisher)

Imidazole (Sigma)

Isopropyl β-D-1-Thiogalactopyranoside (IPTG) (Fisher)

Kanamycin Sulfate (Fisher)

Kodak Storage Phosphor Screen SO230 (Bio-Rad)

Lauryl Sulfate (Research Products International Corp.)

Magnesium Chloride (Fisher)

Maltose (Sigma)

Methanol (Fisher)

Microcon M30 and M50 Centrifugal Filter Devices (Millipore)

MicroSpin G-25 Columns (Amersham Pharmacia Biotech Inc)

MMLV Reverse Transcriptase (Promega)

NiNTA HisPur Resin (Thermo)

Parafilm (Pechiney)

Phosphoric Acid (Sigma)

Poly(A) Polymerase (New England Biolabs)

Poly-Prep Chromatography Columns (Bio-Rad)

Potassium Chloride (Fisher)

Potassium Phosphate Monobasic (Sigma)

Potassium Phosphate Dibasic (Sigma)

Q5 Hot-Start DNA Polymerase (New England Biolabs)

RNasin (New England Biolabs)

Sodium Acetate (Sigma)

Sodium Chloride (Fisher)

Sodium Hydroxide (Sigma)

Sodium Periodate (Sigma)

Sodium Phosphate Dibasic (Fisher)

Succinimidyl Iodoacetate (SIA) (Sigma)

Superscript IV Reverse Transcriptase (Invitrogen)

T4 RNA Ligase I (New England Biolabs)

T4 RNA Ligase II (New England Biolabs)

T7 RNA Polymerase (Epicentre)

Taq DNA Polymerase (Promega)

Tentagel-NH_2 Resin (Chem Impex International)

Tetramethylethylenediamine (TEMED) (Fisher)

Thiol Sepaharose 6B Resin (Sigma)

Tris (Fisher)

TRI Reagent (Sigma)

TRI Reagent (Molecular Research Center Inc.)

Tryptone (Fisher)

UltraLink Iodoacetyl Resin (Thermo)

Urea (Fisher)

Xylene Cyanol (Sigma)

Yeast Extract (Fisher)

3.3 Buffers and Commonly Used Reagents

10X Nudix Buffer, (10 ml, pH 7.5)

Components	10X Concentration	1X Concentration
5 ml 1 M Tris, pH 7.5	500 mM	50 mM
0.5 ml 1 M $MgCl_2$	50 mM	5 mM
4.5 ml Water		

5X Native Phosphate Buffer, NPB (1 L, pH 8)

Components	5X Concentration	1X (NBB)
34.5 g NaH_2PO_4	250 mM	50 mM
146.3 g NaCl	2.5 M	0.5 M
1 L Water, pH 8		

3 M Imidazole Buffer (50 ml, pH 7.6)

Components	Concentration
10.2 g Imidazole	3 M
1.46 g NaCl	50 mM
0.15 g NaH_2PO_4	20 mM
50 ml water, pH 7.6	

Native Wash Buffer, NWB (250 ml, pH 8)

Components	Concentration
50 ml 5X NPB	1X
1.67 ml 3 M Imidazole Buffer	20 mM
Water to 250 ml, pH 8	

Native Elution Buffer, NEB (50 ml, pH 8)

Components	Concentration
10 ml 5X NPB	1X
4.16 ml 3 M Imidazole Buffer	250 mM

Water to 50 ml, pH 8	

1 M Maltose (10 ml)

Components	Concentration
3.6 g Maltose monohydrate	1 M
Water to 10 ml	

NBB+Maltose (50 ml, pH 8)

Components	Concentration
10 ml 5X NPB	1X
1 ml 1 M Maltose	20 mM
Water to 50 ml, pH 8	

1 M $MgCl_2$ (10 ml)

Components	Concentration
2.07 g $MgCl_2$	1 M
10 ml water	

1 M NaCl (50 ml)

Components	Concentration
2.92 g NaCl	1 M
50 ml water	

1 M KCl (50 ml)

Components	Concentration
3.72 g KCl	1 M
50 ml water	

1 M Glucose (50 ml)

Components	Concentration
9 g dextrose	1 M
50 ml water, filter sterilize	

SOC Media (100 ml)

Components	Concentration

2 g tryptone	2%
0.5 g yeast extract	0.5%
1 ml 1 M NaCl	10 mM
0.25 ml KCl	2.5 mM
Water to 100 ml; autoclave	
100 µl 1 M $MgCl_2$	1 mM
200 µl 1 M Glucose	2 mM

LB (1 L)

Components	Concentration
10 g tryptone	10%
5 g yeast extract	5%
10 g NaCl	10 %
Water to 1 L, autoclave	

Phosphate Dialysis Buffer, PDB (1 L, pH 8)

Components	10X Concentration	1X Concentration
27.6 g NaH_2PO_4	200 mM	20 mM
87.63 g NaCl	1.5 M	150 mM
Water to 1 L, pH 8		

Tris/Glycine Running Buffer (2 L)

Components	Concentration
6 g Tris	25 mM
30 g Glycine	200 mM
20 ml 10% SDS	0.1%
Water to 2 L	

10X Tris-Borate-EDTA, TBE (1 L)

Components	10X Concentration	1X Concentration
121.1 g Tris	1 M	100 mM
61.8 g Boric Acid	1 M	100 mM

7.4 g EDTA disodium salt	20 mM	2 mM
Water to 1 L		

4x Laemmli buffer (10 ml)

Components	4X Concentration	1X Concentration
1.6 ml 1.5 M Tris pH 6.8	240 mM	60 mM
0.8 g SDS	8%	2%
3 ml Glycerol	30%	7.5%
0.01 g Bromophenol Blue	0.1%	0.025%
1 ml β-mercaptoethanol	10%	2.5%
Water to 10 ml		

2x TBE Gel Loading Buffer (1 ml)

Components	2X Concentration	1X Concentration
0.2 ml 10x TBE	2X	1X
0.42 g Urea	7 M	3.5 M
0.001 g Bromophenol Blue	0.1%	0.05%

0.001 g Xylene Cyanol	0.1%	0.05%
0.8 ml water		

0.4M KH_2PO_4 (1 L)

Components	Concentration
54.4 g KH_2PO_4	0.4 M
1 L water	

3 M Sodium Acetate (50 ml, pH 5.2)

Components	Concentration
12.3 g sodium acetate	3 M
50 ml water, pH 5.2 with glacial acetic acid	

10% SDS (100 ml)

Components	Concentration
10 g Lauryl sulfate	10%

Water to 100 ml	

2% SDS

Components	Concentration
2 ml 10% SDS	2 %
Water to 10 ml	

1.5 M Tris (1 L, pH 8.8)

Components	Concentration
181.5 g Tris	1.5 M
Water to 1 L, pH 8.8	

0.5 M Tris (50 ml, pH 6.8)

Components	Concentration
3 g Tris	0.5 M
Water to 50 ml, pH 6.8	

8% Denaturing PAGE Solution (1 L)

Components	Concentration
200 ml 40% acrylamide (19:1)	8%
420 g urea	7 M
100 ml 10X TBE	1X
Water to 1 L	

12% SDS-PAGE Gel Solution (15 ml, pH 8.8)

Components	Concentration
3.75 ml 1.5 M Tris pH 8.8	375 mM
0.75 ml 2% SDS	0.1%
6 ml 30% acrylamide (37.5:1)	12%
Water to 15 ml	
20 µl TEMED	
90 ul 10% ammonium persulfate	

4% SDS-PAGE Gel Solution (6 ml, pH 6.8)

Components	Concentration
1.5 ml 0.5 M Tris pH 6.8	125 mM
0.3 ml 2% SDS	0.1%
0.8 ml 30% acrylamide (37.5:1)	4%
Water to 6 ml	
7 µl TEMED	
37 ul 10% ammonium persulfate	

16% SDS-PAGE Gel Solution (15 ml, pH 8.8)

Components	Concentration
3.75 ml 1.5 M Tris pH 8.8	375 mM
0.75 ml 2% SDS	0.1%
8 ml 30% acrylamide (37.5:1)	12%
Water to 15 ml	
20 µl TEMED	
90 ul 10% ammonium persulfate	

3.4 Experimental Procedures

3.4.1 Polymerase Chain Reaction

Polymerase Chain Reaction (PCR) of cDNA, genomic DNA, and geneblock oligos was carried out in 50 µL volumes with 1X Q5 buffer, 0.25 mM dNTPs, 0.5 µM forward and reverse primer, and 0.02 U/µL of Q5 Hot Start DNA Polymerase following the manufacturer's protocol. Resulting products were analyzed by agarose gel electrophoresis and visualized using Molecular Imager FX after staining with ethidium bromide. Plasmids were amplified in 20 µL reactions with 1X Thermophilic DNA Polymerase buffer, 2.5 mM $MgCl_2$, 0.25 mM dNTPs, 0.5 µM primers, and 0.05 U/µL Taq polymerase following the manufacturer's protocol. Products were resolved on agarose gel and visualized similarly.

3.4.2 Preparation of Cloning Fragments

Linearized plasmids were prepared for cloning by PCR using primers to specifically amplify the backbone. The pET-32b vector was prepared by PCR in 50 µL volume with 1X Q5 buffer, 0.25 mM dNTPs, 0.5 µM each of primers 32bF and 32bR, 5 ng pET-32b-GrnA from the Rangachari lab, and 0.02 U/µL Q5. The reaction was cycled between 98 °C for 10 seconds, 60 °C for 10 seconds, and 72 °C for 4 minutes to generate the 5.7 kb linear backbone stock to be used for cloning. The pMBP vector was prepared by PCR in 50 µL volume with 1X Q5 buffer, 0.25 mM dNTPs, 0.5 µM each of PM1 and PM2 primers, 10 ng pMBP-Parallel that was gifted by the Xu lab, and 0.02 U/µL Q5. The reaction was cycled between 98 °C for 10 seconds, 60 °C for 10 seconds, and 72 °C for 3.5 minutes

to generate the 6.7 kb linear vector. The use of the PM1 primer specifically served to replace the embedded TEV protease cleavage site with a thrombin cleavage site for further use.

The NDX8 gene was prepared for insertion into the pETite-nHis-SUMO-Kan vector by 50 µL PCR containing 1X Q5 buffer, 0.25 mM dNTPs, 0.5 µM each of ETitR and ETitSF primers, 12.5 ng of the NDX8 geneblock (IDT), and 0.02 U/µL Q5. The reaction was cycled between 98 °C for 10 seconds, 60 °C for 10 seconds, and 72 °C for 30 seconds to generate the 747 bp NDX fragment with 18 bp overlaps to the pETite-nHis-SUMO-Kan vector that would allow insertion directly downstream of and in frame with the SUMO fusion. The NDX8 gene was prepared for insertion into the pET-32b vector by 50 µL PCR containing 1X Q5 buffer, 0.25 mM dNTPs, 0.5 µM each of primers 32DXF and 32DXR, 12.5 ng of the NDX8 geneblock (IDT), and 0.02 U/µL Q5. The reaction was cycled between 98 °C for 10 seconds, 61 °C for 10 seconds, and 72 °C for 30 seconds to generate the 741 bp NDX8 fragment with 18 bp overlaps to the linearized pET-32b vector that would allow insertion of the gene directly downstream of and in frame with the TrxA fusion.

The Nudt7 gene was prepared for insertion into the pETite-nHis-SUMO-Kan vector by 100 µL PCR containing 5X Q5 buffer, 0.25 mM dNTPs, 0.5 µM each of primers ETitSF7 and ETitR7, 10 ng mouse cDNA from C3H cells from the Guo lab, and 0.02 U/µL Q5. The reaction was cycled between 98 °C for 10 seconds, 61 °C for 10 seconds, and 72 °C for 30 seconds to generate the 747 bp Nudt7 fragment with 18 bp overlaps to the linearized pETite-nHis-SUMO-Kan

vector (Lucigen) that would allow insertion of the gene directly downstream of and in frame with the SUMO fusion.

The NudL gene was prepared for insertion into the pETite-nHis-SUMO-Kan vector by 50 µL PCR containing 1X Q5 buffer, 0.25 mM dNTPs, 0.5 µM each of primers ETitSF6 and ETitR6, 10 ng of *E. coli* genomic DNA, and 0.02 U/µL Q5. The reaction was cycled between 98 °C for 10 seconds, 59 °C for 10 seconds, and 72 °C for 30 seconds to generate the 615 bp NudL fragment with 18 bp overlaps with the vector to allow insertion downstream of and in frame with the SUMO fusion. The NudL gene was prepared for insertion into the pMBP-Parallel vector by 50 µL PCR containing 1X Q5 buffer, 0.25 mM dNTPs, 0.5 µM each of primers MNF1 and MNR1, 5 ng of pETS-NudL, and 0.02 U/µL Q5. The reaction was cycled between 98 °C for 10 seconds, 60 °C for 10 seconds, and 72 °C for 30 seconds to generate the 618 bp NudL fragment with 18 bp overlaps to the linearized vector that would allow insertion directly downstream of and in frame with the maltose binding protein fusion. The MNF1 primer specifically was used to insert the thrombin cleavage site in place of the TEV cleavage site embedded in the backbone before its preparation, and therefore the addition of this thrombin site by MNF1 upstream of NudL and the same by PM1 downstream of the insertion site in pMBP-Parallel served as one site of recombination.

The NudC gene was prepared for insertion into the pMBP-Parallel vector by 50 µL PCR containing 1X Q5 buffer, 0.25 mM dNTPs, 0.5 µM each of primers MNF2 and MNR2, 10 ng of *E. coli* genomic DNA, and 0.02 U/µL Q5. The reaction was cycled between 98 °C for 10 seconds, 50 °C for 10 seconds, and 72 °C for

45 seconds to generate the 813 bp NudC fragment with 18 bp overlaps to the linearized vector that would allow insertion directly downstream of and in frame with the maltose binding protein fusion. The MNF2 primer specifically was used to insert the thrombin cleavage site in place of the TEV cleavage site embedded in the backbone before its preparation. The NudC gene was prepared for insertion into the pETite-nHis-SUMO-Kan vector by 50 µL PCR containing 1X Q5 buffer, 0.25 mM dNTPs, 0.5 µM each of primers ETitSF8 and ETitR8, 25 pg of PCR product used to generate the pEMS vector, and 0.02 U/µL Q5. The reaction was cycled between 98 °C for 10 seconds, 50 °C for 10 seconds, and 72 °C for 40 seconds to generate the 810 bp NudC fragment with 18 bp overlaps with the vector to allow insertion downstream of and in frame with the SUMO fusion.

3.4.3 Transformation

Following the protocol by Lucigen, competent subcloning cells (High-Control 10G, DH5α) were transformed by the addition of 12.5 ng prepared linearized plasmid and 12.5 ng prepared plasmid insert. Typical transformation protocols followed, where cells were incubated with DNA on ice for 30 minutes, heat shock for 30 to 45 seconds, recovery on ice for 5 minutes, and then addition of 500 µL SOC for 1 hour outgrowth at 37 °C shaking before dilution and plating on antibiotic solid media for overnight incubation at 37 °C.

3.4.4 Colony PCR, Plasmid Analysis, and Culture Handling

All media used to grow bacteria were LB or LB agar supplemented with the appropriate antibiotic dictated by the plasmid used. Transformants were picked as isolated colonies from antibiotic solid media by a pipet tip, smeared

into a 0.2 µL PCR tube, and dipped into 1 mL antibiotic media that was then incubated at 37 °C to propagate the culture. Batch mixtures for PCR were used that contained 1X Thermophilic DNA Polymerase buffer, 2.5 mM MgCl2, 0.25 mM dNTPs, 0.5 µM primer pairs, and 0.05 U/µL Taq polymerase. The batch mixture was added to each 0.2 µL tube as 18 µL aliquots, and the samples were heated at 95 °C for 10 minutes to lyse the cells before cycling between 95 °C for 45 seconds, the annealing temperature for 45 seconds, and 72 °C for 1 minute to extend each target considering they all were under 1 kb in size. General backbone primers for pETite-nHis-SUMO-Kan constructs were SUMOFwd and pETiteRev with the annealing temperature of 51 °C, and PM3 and PM4 with an annealing temperature of 50 °C for the pMBP-Parallel vector. Products were resolved on agarose gel, and samples with positive results were removed from 37 °C at the first signs of turbidity where aliquots were frozen with 40% glycerol at -80 °C.

The freezer stocks were used to inoculate 5 mL antibiotic media for overnight growth at 37 °C shaking, and the following day plasmid extraction was carried out using the Epoch Miniprep kit following the manufacturer's protocol. DNA quantity was measured by absorbance at 260 nm, aliquots of samples were diluted to 1 ng/µL and used as templates for 20 µL PCR containing 1X Thermophilic DNA Polymerase buffer, 2.5 mM $MgCl_2$, 0.25 mM dNTPs, 0.5 µM appropriate primers, 1 ng plasmid DNA, and 0.05 U/µL Taq polymerase. The reactions were cycled between 95 °C for 45 seconds, annealing temperature for

45 seconds, and 72 °C for 1 minute. Products were visualized on agarose gel to qualify successful generation of target plasmid.

3.4.5 Small Scale Protein Expression and Solubility Assay

Transformation into expression cell lines (High-Control Bl21 DE3, Rosetta 2 DE3, NEBExpress) following the manufacturer's protocol was carried out following plasmid verification by PCR. Cultures were analyzed by Colony PCR in the same way as with subcloning cell lines, where transformants producing positive results from Colony PCR were frozen in 40% glycerol at -80 °C, while 100 µL of the 1 mL antibiotic media grown to turbidity was used to inoculate 5 mL antibiotic media that was incubated at 37 °C. When the absorbance at 600 nm reached 0.5, 1 mL of the culture was removed, pelleted, resuspended in 50 µL 4X Laemmli buffer and stored at -20 °C (uninduced lysate), and the remaining culture was induced with 0.2 to 0.5 mM IPTG to express either overnight at 16 °C or for 2 to 3 hours at 37 °C. After expression, absorbance at 600 nm was obtained, 1 mL culture was removed to be pelleted and resuspended in 100 µL 4X Laemmli buffer and saved at -20 °C (induced lysate). Before running on SDS-PAGE, small scale cell lysate samples were lysed at 95 °C for 5 minutes and then spun at 5,000 x g for 5 minutes. The remaining culture was pelleted and resuspended in 1 mL NBB for lysis either by heating at 95 °C for 10 minutes or sonicating on ice for 5 cycles of 15 seconds on with 45 seconds rest using a probe sonicator with the XL-2000 Misonix Ultra Liquid Processor on power 18. The resulting cell lysate was pelleted by spinning 16,000 x g for 15 minutes, and

after removing the supernatant (soluble fraction) the pellet was resuspended in another 1 mL NBB (insoluble fraction).

3.4.6 Large Scale Protein Expression and Cell Lysate Preparation

Upon determination of positive protein expression and solubility, freezer stocks were used to inoculate 5 mL antibiotic media for overnight growth at 37 °C shaking. The following day, the overnight culture was used to inoculate 0.5 L of antibiotic media and was grown until absorbance at 600 nm reached 0.5, where 1 mL was then removed to be pelleted and resuspended in 50 µL 4X Laemmli buffer (uninduced lysate as before), and the remaining culture was induced with 0.2 to 0.5 mM IPTG for expression at 16 °C overnight shaking. The following day, 1 mL culture was removed to obtain absorbance at 600 nm, and then was pelleted and resuspended in 100 µL 4X Laemmli buffer (induced cell lysate). The remaining culture was pelleted at 10,000 x g for 10 minutes at 4 °C in an Avanti J-26 XP centrifuge using a JLA 16.250 rotor and frozen at -20 °C for at least 1 hour. Cells were resuspended in 8 mL NBB (pH 8) with an 18-G syringe and sonicated on ice with the XL-2000 sonicator with probe on power 18 for 5 cycles of 15 seconds on and 45 seconds resting. The cell lysate was clarified by spinning at 40,000 x g for 30 minutes at 4 °C in an Avanti J-26 XP centrifuge with a JA 25.50 rotor, and the supernatant was taken as the soluble protein fraction. The resulting pellet was resuspended in 8 mL NBB (pH 8) and considered the insoluble protein fraction for SDS-PAGE.

3.4.7 Protein Purification

Proteins were purified by either amylose resin for MBP-fusions or IMAC using NiNTA resin for His-tagged proteins. Amylose resin was prepared by addition of 1 mL of 50% ethanol slurry to Poly-Prep Chromatography column and allowing to settle to approximately 0.5 mL. Storage solution was removed and column was washed with 10 column volumes (cv) of water, 10 cv 0.1% SDS, and 20 cv water before equilibrating with 10 cv NBB (pH 8). Soluble cell lysate prepared as stated was added and passed over the resin three times while collecting the flow-through. The column was washed with NBB until absorbance at 280 nm was zero (approximately 30 cv), and the protein was eluted with the addition of 4 cv NBB + 20 mM maltose. The absorbance at 280 nm of each fraction was obtained and they were visualized by SDS-PAGE. Eluted fractions containing purified protein were combined on an M50 spin column to reduce the volume to approximately 300 µL, and the combined protein was desalted by the addition of equal volumes 1X PDB. Samples were supplemented with 40% glycerol before absorbance at 280 nm was taken and final concentration was calculated using the extinction coefficient of individual proteins before storage at -20 °C. The final product was run on SDS-PAGE to visualize purity.

Proteins purified by NiNTA resin followed a similar protocol with some changes. NiNTA resin was prepared by addition of 1 mL 50% ethanol slurry to a Poly-Prep Chromatograph Column and allowed to settle to approximately 0.5 mL, washed with 10 cv water, and equilibrated with 10 cv NBB (pH 8). Soluble cell lysate was passed over the resin 3 times while collecting flow-through, and

column was washed with 8 cv NBB or until absorbance at 280 nm reached zero, followed by 8 cv NWB (containing 20 mM imidazole) or until absorbance at 260 nm reached zero. The protein was eluted with 8 cv NEB (containing 250 mM imidazole) or until absorbance at 280 nm was zero. Samples were visualized on SDS-PAGE, and elution fractions were combined on M30 column to reduce the volume to approximately 300 µL where they were desalted by the addition of equal volumes of 1X PDB. Protein was supplemented with 40% glycerol before taking absorbance at 280 nm and calculating final concentration using the extinction coefficient of each protein and storing at -20 °C.

3.4.8 Enzyme Activity Assays

Enzyme activities were tested by incubation of enzyme (Nudt7, NudC, or NudL) in 1X Nudix buffer in the presence of various concentrations of either CoA, FAD, or NAD+ at 37 °C in an Eppendorf Mastercycler. Reaction progress was monitored by HPLC using a 250 x 4.6 mm Allspher SAX column with an isocratic elution protocol of 1.5 mL/min with 60% KH_2PO_4. Products and reactants were resolved into separate peaks which were integrated to determine the respective areas using Origin 8.5 software, which was also used to determine Km, Vmax, and Kcat upon plotting rate vs. substrate concentration.

Nudt7 activity against CoA thioesters were carried out by the combination of 16 µM Nudt7 in 1X Nudix buffer with 120 pmol internally labeled CoA-RNA samples before and after biotinylation. The biotinylation was carried out by incubating 60 pmol labeled CoA- RNA (with and without Nudt7 treatment) with 19 nmol Biotin-AMP in the presence of 500 mmol imidazole at room temperature for

30 minutes. The reaction was precipitated with the addition of 0.3 M NaAc (pH 5.2) and 3:1 ethanol at -20 °C for 30 minutes, and the RNA was recovered in water for treatment with or without Nudt7 followed by another ethanol precipitation. Reactions were combined with 2 µg streptavidin at room temperature for 10 minutes and then run on 8% denaturing PAGE to resolve the RNA and Streptavidin-Biotin-CoA complex, where the gels were dried in a Model 853 gel dryer and exposed to Kodak Storage Phosphor Screen and visualized using a PhosphorImager to observe the shift in electrophoretic mobility.

3.4.9 RNA and CoA-RNA Preparation

RNA was prepared by In Vitro Transcription of dsDNA templates bearing the φ2.5 promoter with the Ampliscribe T7 Flash kit from Epicentre following the manufacturer's protocols and previously described methods (Huang 2003b). Briefly, reactions of 10 µL contained the manufacturer specified components (1X Reaction buffer, 2.5 mM NTPs, 10 mM DTT, 1 U RNasIn, 1X dsDNA template, and 1 µL of T7 RNA polymerase) and were incubated at 37 °C for 1 to 3 hours. Short RNA under 100 nt was desalted by gel filtration with MicroSpin G-25 Columns, precipitated with 0.3 M NaAc (pH 5.2) and 3:1 ethanol at -20 °C for 30 minutes, and then purified by 8% denaturing PAGE. Isolated bands on PAGE were visualized by UV shadowing over a TLC plate and excised. RNA was extracted from gel in the presence of 0.3 M NaAc under heat and precipitated with the addition of 3:1 ethanol at -20 °C for 30 minutes. RNA was recovered in water and quantified by absorbance at 260 nm. RNA longer than 100 nt were desalted and purified by M30 centifugal filter, followed by ethanol precipitation

and quantification similarly. CoA-RNA was prepared similarly by IVT except with the addition of 10 mM depCoA reaction mixture, and purification and desalting was carried out by M30 centrifugal filter followed by precipitation. CoA-RNA was purified from pppRNA by Thiol Sepharose 6B resin (Sigma) by binding in the presence of 50 mM Tris (pH 8) for multiple passes, extensive washing with excess 3 M NaCl, 4 M NaAc (pH 5.2), 4 M Urea, and water, followed by three rounds of 3 M NaCl and water, and a final round of 3 M NaCl, 4 M NaAc (pH 5.2), 4M Urea, and water. CoA-RNA was eluted with 0.5 mL 100 mM DTT in 100 mM Tris (pH) and 10 µg glycogen as a carrier, and the column was washed with 400 µL 0.4 M NaAc (pH 5.2) and 10 µg glycogen. The collected flow-through was precipitated by the addition of 3:1 ethanol at -20 °C, and the RNA was recovered in pure water.

Internally labeled CoA-RNA was prepared by IVT similar to CoA-RNA above, except with the addition of 1:2 NTP mixture (containing half the amount of ATP as CTP, UTP, and GTP) instead of the NTP mixture, as well as the addition of 5 µCi α-^{32}P-ATP. RNA was desalted and purified by M30 centrifugal filter and precipitated as described above. CoA-RNA was alternatively internally labeled by ligation with a 5'-labeled RNA adapter. Adapters were 5'-labeled by reacting 100 pmol RNA with 30 µCi γ-^{32}P-ATP and T4 Polynucleotide Kinase (T4PNK) in 1X T4PNK buffer for 30 minutes at 37 °C. The CoA-RNA was combined with labeled adapter at various molar ratios in 1X T4Rnl1 buffer with 1 mM ATP and 0.25 U/µL T4 RNA Ligase 1 (T4Rnl1) for 1 h at 37 °C. Ligation reaction efficiency was visualized by 8% denaturing PAGE stained with ethidium bromide for

visualization by the Molecular Imager FX if RNA was not labeled, or dried and exposed to a Kodak Storage Phosphor Screen for visualization by the PhosphorImager if RNA was labeled.

3.4.10 CoA Conjugation to Resin

CoA and CoA-RNA were conjugated to resins functionalized with a terminal iodo group (Tengtagel-NH2-SIA, UltraLink Iodoacetyl, UltraLink PC-Linker Iodo, and ULPCIA) by incubation in 10-50 mM Tris (pH 8) at room temperature while vortexing. The capacity and time of reaction for each resin was assessed by room temperature incubation in the presence of 10 mM Tris (pH 8) and 0.8 mM β-mercaptoethanol while quantifying the release of free iodide with the absorbance at 226 nm and the experimentally determined extinction coefficient of 14,000 M-1 cm-1. The total amount of iodide released was considered the capacity of the resin. After reacting CoA or CoA-RNA with each resin, the remaining functional groups were blocked by similar incubation upon the addition of excess β-mercaptoethanol. The resin was subjected to extensive washing to remove non-covalently interacting CoA or CoA-RNA with a total of 40 bed volumes alternating between water, 4 M NaCl, and 8 M urea while heating briefly at 95 °C and vortexing.

3.4.11 CoA Release Assays

CoA release by enzyme hydrolysis was carried out in the presence of 1X Nudix buffer with 16 μM Nudt7 at 37 °C for 30 minutes, 3 hours, or overnight incubation. Reaction progress was determined by the absorbance at 260 nm of the 3’5’-ADP product of CoA hydrolysis in the reaction solution. After removal of

reaction solution, beads were washed with 10 bed volumes of water to release any CoA stuck in the beads due to settling for similar quantification. The release of radiolabeled CoA-RNA was carried out in a similar manner, however the released RNA was assessed by the determination of the radioactivity in the reaction solution as well as remaining on the resin by liquid scintillation counter. The release of CoA-RNA by resins containing photocleavable linkers was achieved by the exposure to 365 nm light by the 100 W Blak-Ray UV lamp for 5 minutes at a time on ice with the beads suspended in water. CoA release was determined by absorbance of reaction solution at 260 nm of the CoA-thioester product, as well as by electrospray ionization mass spectrometry (ESI-MS) on negative mode.

3.4.12 RNA Ligation and Reverse Transcription

RNA Ligation was optimized by reactions containing 1X T4Rnl1 buffer, 1 mM ATP, and 0.2 U/μL T4 RNA Ligase 1 incubated at 37 °C for 30 minutes to 1 hour. RNA containing 5’ppp was generated by In Vitro Transcription as described above. RNA containing 5’p was produced by IVT as described above with the addition of 4 molar excess AMP, and its 3’ end was subsequently blocked from potential ligation with oxidation by 10 mM sodium m-periodate for 40 minutes on ice in the dark followed by ethanol precipitation. Synthetic RNA oligos containing a 5’OH were prepared for ligation by phosphorylation with 1X T4PNK buffer, 1 mM ATP, and 0.2 U/μL T4PNK at 37 °C for 30 minutes. Adenylated DNA was produced by the addition of 30 mM MgCl2 and 120 mM adenosine phosphoimidazolide (ImpA) to 5’pDNA oligos at 50 °C for 3 hours, followed by

purification with Amicon 3K spin column, ethanol precipitation, and purification by 10% denaturing PAGE. The addition of up to 15% DMSO or PEG8000 (v/v) in ligation reactions was noted to increase the ligation efficiency in terms of amount of product visualized by PAGE.

Reverse Transcription was initially carried out in 20 µL reactions composed of 15 µL solutions containing 2 µL RNA, 2 µM RT primer, and 0.25 mM dNTPs that were heated for 5 minutes at 70 °C to denature RNA, cooled on ice to allow annealing of primer, and extended at 42 °C for one hour after the addition of 4 µL 5X MMLV buffer and 200 U MMLV RTase. Later experiments with the more thermostable SuperScript IV enzyme were carried out in a similar way with the exception the addition of 5 mM DTT, 1X SSIV buffer, and 1.5 µL SSIV occurring either with or without cooling on ice. Extension of cDNA was carried out at temperatures varying from 37 °C to 70 °C followed by the optional denaturation of the enzyme for 5 at 95 °C.

3.4.13 CoA-RNA Capture Protocol

Total RNA from E. coli was prepared by growing High Control 10G cells transformed with pETite-SUMO-nHis-Kan vector encoding the citrate binding protein CitAP in 100 mL LB media with 50 µg/mL kanamycin overnight at 37 °C shaking. Cells were pelleted at 5,000 x g and resuspended in TRI Reagent for total RNA extraction following the manufacturer's protocol. Total RNA was resuspended in water, quantified by absorbance at 260 nm, and then treated with 100 mM DTT for 30 minutes at room temperature followed by ethanol precipitation twice to isolate the RNA. The reduced RNA was resuspended in

water and transferred to 50 µL Thiol Sepharose 6B resin where it passed over the resin in the presence of 70 mM Tris (pH 8) ten times (Column A), and the flow-through of Column A was kept. The column was then extensively washed to remove nonspecifically interacting RNA by the addition of excess 3 M NaCl, 4 M NaAc (pH 5.2), 4 M urea, and water for one round, followed by three rounds of excess 3 M NaCl and water, and a final round of 3 M NaCl, 4 M NaAc (pH 5.2), 4 M urea, and water. The column was then treated with 0.5 mL of 100 mM DTT in 100 mM Tris (pH 8) with 10 µg glycogen to release the RNA (Sample A) from the column, and the initial flow-through was collected by addition of 0.4 mL 0.4 M NaAc (pH 5.2) and 10 µg glycogen so it could be precipitated with the addition of ethanol.

Flow-through from Column A was precipitated with 0.4 M NaAc (pH 5.2) and ethanol, and resuspended in water for treatment with 50 mM hydroxylamine and 20 mM DTT for 15 minutes at room temperature to convert thioester-CoA-RNA to CoA-RNA, and after ethanol precipitation was added to a secondary Thiol Sepharose resin (Column B) to be purified similar to above. The resulting RNA (Sample B) was considered to be CoA-RNA not captured from Column A as well as thioester-CoA-RNA. The initial flow-through of Column B was collected, ethanol precipitated, and purified by a tertiary Thiol Sepharose resin (Column C) similar to above, where the purified RNA (Sample C) was considered to be a background control. The purified RNA from Columns A, B, and C were subjected to secondary columns each (Columns A', B', and C') following a similar protocol. The recovered RNA was then treated with 16 µM Nudt7 in 1X T4Rnl1 buffer for 1

hour at 37 °C to remove the CoA and leave a 5'monophosphate, where it was subsequently ligated to the 5' adapter GL1a in the presence of 1 mM ATP, 0.2 U/µL T4 RNA Ligase 1, 1 µL RNasIn, and 15% DMSO at 37 °C for 1 hour. Reverse transcription was carried out by the addition of 2 µL RNA to 5 µL RT Stock A (2 µM Adp16 and 2.5 mM dNTPs). The reaction mixture was heated for 5 minutes to 70 °C, and after the addition of 3 µL Stock B (1X SSIV buffer, 5 mM DTT, 1.5 µL SSIV) the reaction was incubated for 10 minutes at 50 °C. After RT, the enzyme was denatured by incubation at 95 °C for 5 minutes. The resulting cDNA was amplified by PCR in 10 µL with 1X Q5 buffer, 0.25 mM dNTPs, 0.6 µM Adp2, 0.6 µM PcR2 and 0.02 U/µL Q5 for 30 cycles of 98 °C for 20 seconds, 57 °C for 20 seconds, and 72 °C for 60 seconds. Results were visualized by agarose gel.

3.5 DNA and RNA Oligos

Table 3.1

DNA and RNA Oligos

Name	5'	Sequence	3'
Adp2	OH	CGACTCTAGAGGATCCCCTTGGCTTCCATTTCTTTGA	OH
Adp16	OH	TGGCACCCGAGAATTCCAT$_{(28)}$VN	OH
Adp21	OH	CGACTCTAGAGGATCCCCGCATGGCTGACTAGG	OH

Adp26	P	GTTCAGAGTTCTACAGTCCGArGrGrG	OH
Bap12	OH	ATGGTGAGCAAGGGC	OH
SP10	OH	GTGGCGGCCGCTCTATTACTTGTACAGCT CGTCC	OH
PcR2	OH	TGGCACCCGAGAATTCCA	OH
GL1a	OH	UUGGCUUCCAUUUCUUUGAUU	OH
T7Ad	PPP	AGGUACGCUCCGAAA	OH
Ad3	P	CCTAGTCAGCCATGC	NH_2
ETitF	OH	CATCATCACCACCATCACATGAAGTGTGTG GTTAGCC	OH
ETitR	OH	GTGGCGGCCGCTCTATTACTAAATTTTGGA TTTCGGTTCAAA	OH
ETitF1	OH	ATGAAGTGTGTGGTTAGCC	OH
ETitR1	OH	CTAAATTTTGGATTTCGGTTCAAA	OH
ETitSF	OH	CGCGAACAGATTGGAGGTATGAAGTGTGT GGTTAGCC	OH
32DXR	OH	GGTGGCAGCAGCCAATCACTAAATTTTGG	OH

		ATTTCGGTTCAAA	
32DXF	OH	CTGGTGCCACGCGGTTCTATGAAGTGTGTGGTTAGCC	OH
32bF1	OH	CATAGAACCGCGTGG	OH
32bR1	OH	TGATTGGCTGCTGCC	OH
ETitSF7	OH	CGCGAACAGATTGGAGGTATGTCGCGACCTTGTG	OH
ETitR7	OH	GTGGCGGCCGCTCTATTATCACAACTTGCTTAAAGAATATCTC	OH
ETitSF8	OH	CGCGAACAGATTGGAGGTATGGATCGTATAATTGA	OH
ETitR8	OH	GTGGCGGCCGCTCTATTATCACTCATACTCTGC	OH
ETitSF6	OH	CGCGAACAGATTGGAGGTATGGAATACCGTAGCCTG	OH
ETitR6	OH	GTGGCGGCCGCTCTATTATCAGGTTTCACACCAAT	OH
SUMOFwd	OH	ATTCAAGCTGATCAGACCCCTGAA	OH

pETiteRev	OH	CTCAAGACCCGTTTAGAGGC	OH
pM1	OH	AGAACCGCGTGGCACCAGGGTCGTATCGATCCCGAGG	OH
pM2	OH	GCCATGGATCCGGAATTC	OH
pM3	OH	CCTCGGGATCGATACG	OH
pM4	OH	GAATTCCGGATCCATGGC	OH
MNR2	OH	GAATTCCGGATCCATGGCTTATCACTCATACTCTGC	OH
MNF2	OH	CTGGTGCCACGCGGTTCTATGGATCGTATAATTGA	OH
MNF1	OH	CTGGTGCCACGCGGTTCTATGGAATACCGTAGCCTG	OH
MNR1	OH	GAATTCCGGATCCATGGCTTATCAGGGTTTCACACCAATTTGC	OH

All sequences listed are DNA with the exception of GL1a and T7Ad. In the event where oligos are DNA/RNA hybrids, RNA is denoted by lower case r in front of the nucleotide. Cloning primers contain underlined sequences that represent overlap with the gene of interest, where the remainder of the sequence represents overlap with the vector used.

3.6 Cell Lines and Genotypes

Bl21 Competent E. coli Cells (New England Biolabs) - fhuA2 [lon] ompT gal [dcm] ΔhsdS

Bl21 (DE3) Competent E. coli Cells (New England Biolabs) - fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5

High-Control 10G Competent *E. coli* Cells (Lucigen) – *mcrA* Δ (*mrr-hsdRMS-mcrBC) end*A1 *rec*A1 φ80d*lacZ*ΔM15 Δ*lac*X74 *ara*D139 Δ(ara,leu)7697 *gal*U *gal*K *rps*L (Str^{R}) *nup*G λ^{-} *ton*A /Mini-F *lacI^{q1}* (Gent^{R})

High-Control Bl21 (DE3) Competent *E. coli* Cells (Lucigen) – F^{-} *ompT hsdSB (rB- mB-) gal dcm* (DE3)/Mini-F *lacI^{q1}* (Gent^{R})

NEB5-alpha Competent E. coli Cells (New England Biolabs) - Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- ϕ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1 Δ(mrr-hsdRMS-mcrBC)

NEBExpress Competent E. coli Cells (New England Biolabs) - fhuA2 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 Δ(mcrC-mrr)114::IS10

Rosetta 2 (DE3) Competent *E. coli* Cells (Novagen) - F^{-} *ompT hsdSB (rB- mB-) gal dcm* (DE3) pRARE2 (Cam^{R})

CHAPTER IV – THE CLONING AND ACTIVITY OF NUDIX COA PYROPHOSPHOHYDROLASES

4.1 Introduction

The Nudix superfamily of enzymes is largely known for the hydrolyzing activity of its members (McLennan 2006). With members characterized from a variety of different species, the majority of these enzymes target nucleotide derivatives (McLennan 2006, Kupke et al. 2009). Due to commonalities of identified substrates, early characterizations of the Nudix superfamily members established their antimutator potential and function in housecleaning (Bessman et al. 1996). Further evaluation of the superfamily resulted in the identification of members capable of hydrolyzing coenzymes, namely Coenzyme A, and representatives of such a CoA-hydrolyzing role were found in bacteria, yeast, worms, mammals, and plants (McLennan 2006, Cartwright et al. 2000, Gasmi and McLennan 2001, AbdelRaheim and McLennan 2002, Kupke et al. 2009). Such a demonstration of Coenzyme A hydrolysis is intriguing because despite being well-characterized in the 1980s, the portion of the CoA synthetic pathway responsible for CoA degradation had never been address in the model species of *E. coli* (Jackowski and Rock 1981). While most of the enzymes in this pathway are capable of catalyzing their respective reverse reactions, such an activity from CoaD or Phosphopantetheine Adenylyl Transferase (PPAT) requires dephospo-CoA, which is not known to exist in any significant concentration in the cell, and is distinctly different from mature CoA that is so prevalent. The Nudix CoA pyprophosphohydrolases, or CoAses as they

became known, however, are capable of degrading CoA into 3'5'-ADP and 4'-phosphopantetheine.

The discovery of Nudix CoAse enzymes became particularly important upon the realization that bacterial RNA can be capped at its 5' end with CoA (Kowtoniuk et al. 2009). Being the first indication of a bacterial capped RNA more complicated than the incidental 5'-triphosphate group, the CoA cap as well as the concurrently discovered NAD-cap have become the focus of investigation to uncover their function and the identity of the associated sequences (Chen et al. 2009). While examples of NAD-RNA sequences have since been identified using a specific capture chemistry for NAD^+, there has been no capture protocol comparably specific for CoA, therefore leaving these RNAs a mystery (Cahova et al. 2015). We believe that the Nudix CoAse enzymes are the key to the identification of CoA-RNA sequences.

The cloning and recombinant expression of such an enzyme is vital to a specific capture protocol to be established for the isolation of CoA-RNA and the identification of its sequences. Several Nudix CoAses have the potential for use in a CoA-releasing protocol, such as NDX8 from *C. elegans* that shows the highest enzyme efficiency of all characterized CoAses, and Nudt7 from *M. musculus* that shows a high activity towards CoA thioesters (AbdelRaheim and McLennan 2002, Gasmi and McLennan 2001). In addition to the identification of RNA sequences, studying bacterial Nudix enzymes may shed light on the unknown mechanism of CoA-RNA degradation. Investigators have shown the enzyme NudC from *E. coli* to be active against NADH, NAD^+, and CoA, however

this enzyme has been previously uncharacterized for the latter CoA (Bird et al. 2016, Frick and Bessman 1995, Xu et al. 2000). Their demonstration that NAD and CoA can be hydrolyzed by the same enzyme indicates a lack of substrate specificity based on the distinct structural differences of nicotinamide and phosphopantetheine. Indeed, the initial characterization of NudC similarly showed its capability to hydrolyze a variety of nucleotide-derived substrates, however if such different coenzymes can be similarly hydrolyzed, there is a high probability that FAD can be similarly digested by NudC, and therefore NudC can be demonstrated as a general coenzyme hydrolase in *E. coli (Frick and Bessman* 1995*)*. On the other hand, previous investigators have predicted the *E. coli* CoAse to be NudL (also known as YeaB) based on identification of an open reading frame encoding a Nudix box consensus sequence common to all Nudix enzymes that is also downstream of a CoA binding consensus sequence common to all characterized Nudix CoAses (McLennan 2006). Therefore, the cloning, purification, and characterization of Nudt7, NudC, and NudL are reported here with the goal of establishing NudC as a general coenzyme hydrolase, NudL as the *E. coli* CoAse, and Nudt7 as a specific Nudix CoAse capable for use in a CoA-RNA capture protocol.

4.2 Results and Discussion

4.2.1 C. Elegans NDX8

4.2.1.1 Cloning of NDX8

The cloning of NDX8 from *C. elegans* was carried out with the goal of amassing a stock of high efficiency enzyme that can be used for a specific CoA-

RNA capture protocol. As this organism had been sequenced, the most direct route to obtaining the enzyme was determined to be ligation independent cloning of a long DNA oligo into a linear vector following portions of a protocol obtained from Lucigen cloning systems. Considering that DNA ligation is not required, at least one day was removed from the timeline of typical plasmid preparation protocols (Fig. 4.1). Plasmids are constructed *in vivo* by a homologous recombination event that occurs during transformation facilitated by 18 bp overlap regions of the vector and insert. The 18 bp overlap size was determined through different experiments that each shortened the overlap until the resulting number of transformants dramatically decreased. This recombination event is not exclusive to the cells obtained from Lucigen, as this method has been shown to be successful in the common subcloning cell line DH5α as well as expression cell lines Bl21(DE3), Bl21, and Rosetta 2 (DE3) cells that already contain a plasmid. By designing the primers to amplify the insert and linearize the vector, the generation of a plasmid can be completed within 24 hours including the overnight incubation of competent cells on selective antibiotic media.

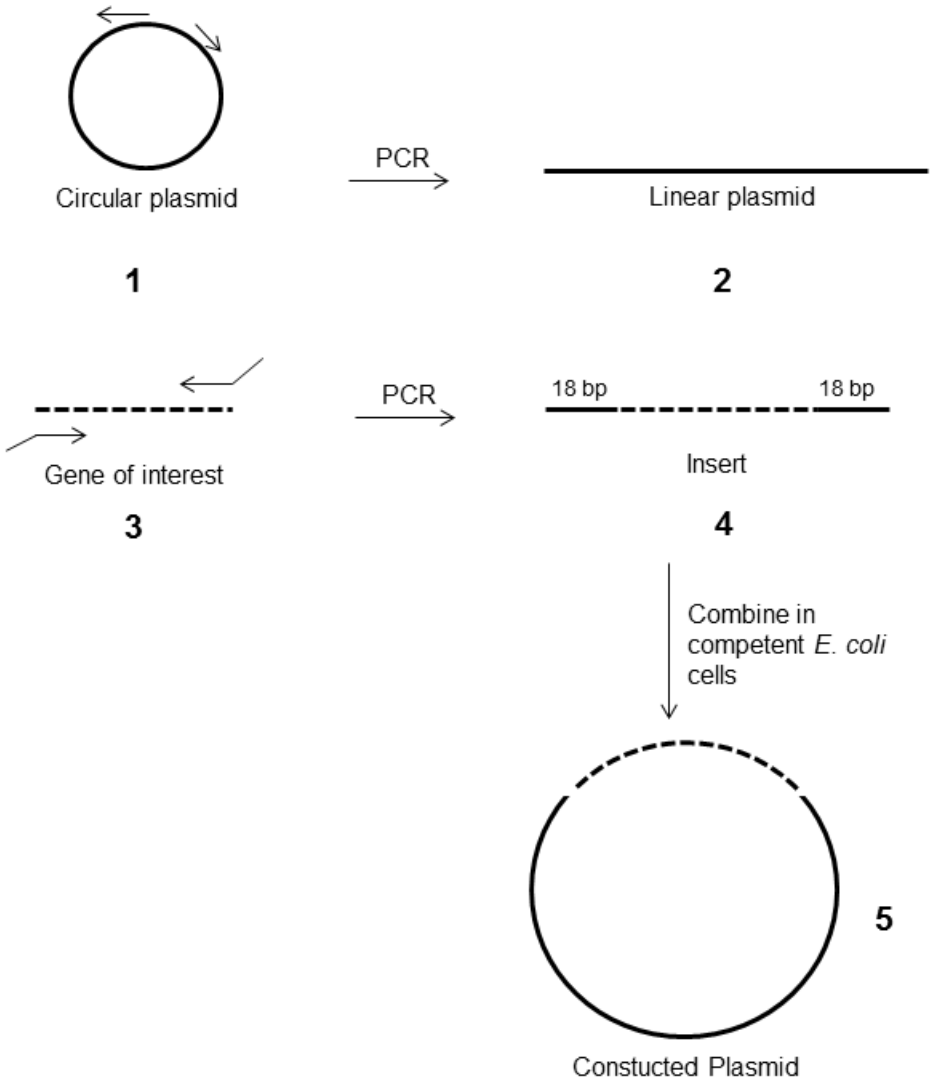


Figure 4.1 Ligation Independent Cloning.

Preparation of plasmids (1) and inserts (2) is carried out by PCR to linearize the vector (3) and to add 18 bp homologous regions flanking the gene of interest (4). When combined in competent *E. coli* cells, the two fragments recombine to form the desired plasmid (5).

The NDX8 gene was synthesized as a 705 bp geneblock by IDT to contain flanking 18 bp regions homologous to the cloning site of the pETite-nHis-Kan cloning vector from Lucigen. The synthesized geneblock and linearized vector

were combined in in High-Control 10G *E. coli* cells for transformation, and the next day colony PCR confirmed successful recombination into 90% of screened colonies to generate the pET-NDX8 plasmid (Fig. 4.2A). One-third of the successful colony PCR reactions showed a significantly intense band near the 766 bp marker. The amplification efficiency of the 705 bp fragment may vary depending on a number of factors, one example of which is the amount of bacteria used in the analysis.

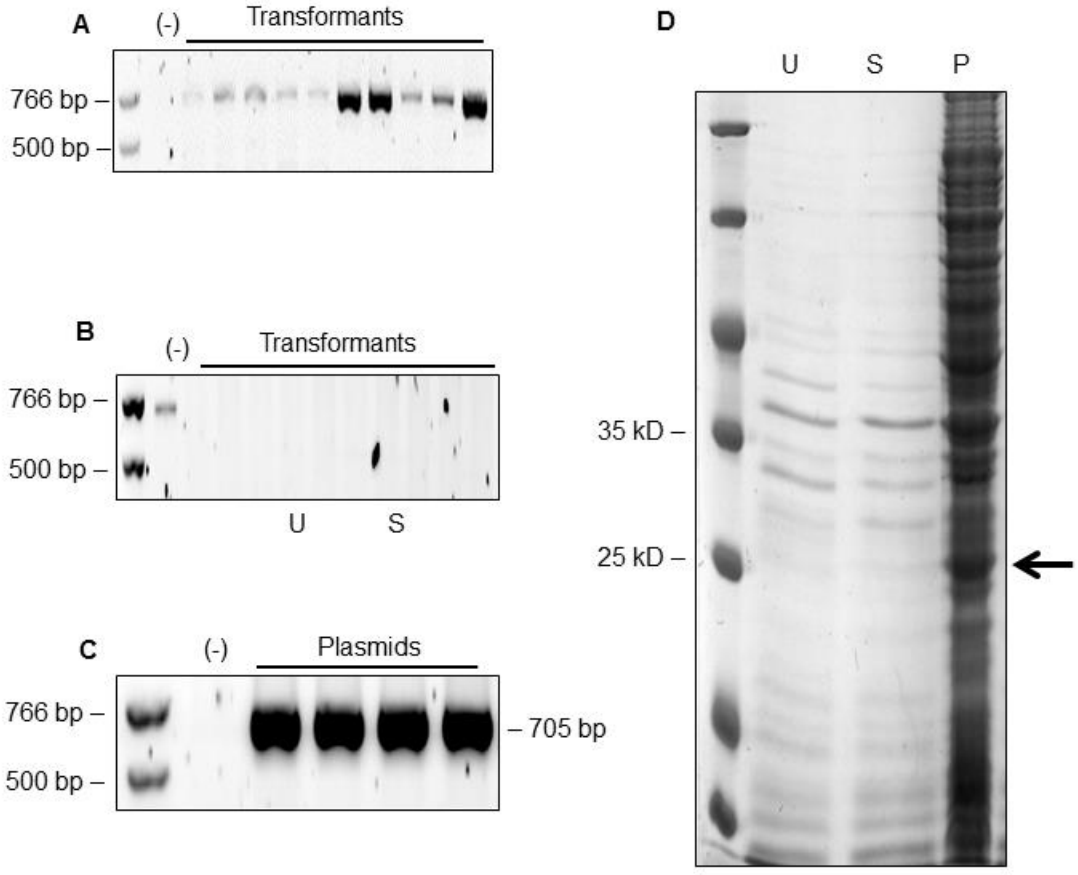


Figure 4.2 Cloning of NDX8 with pETite-nHis-Kan vector

A, Colony PCR results of transformation into High-Control 10G cells using general backbone primers to amplify the 705 bp NDX8 insert. B, Colony PCR results of transformation into Bl21 (DE3) cells using general backbone primers to amplify the NDX8 insert. C, PCR analysis of extracted plasmids from Bl21 (DE3) cultures with failed Colony PCR. D, SDS-PAGE of uninduced (U), soluble (S), and pellet (P) fractions of small scale expression in Bl21 (DE3) cells.

Plasmids were isolated from cultures with positive colony PCR results and used to transform High-Control Bl21 (DE3) cells from Lucigen, whereupon

Colony PCR analysis of transformants showed 100% negative results (Fig. 4.2B). Considering the transformed cells grew in media supplemented with kanamycin, a failed transformation was highly unlikely since kanamycin resistance is granted by the successful uptake of the pETite plasmid. To confirm the problem was with Colony PCR and not transformation, samples from which Colony PCR was performed were grown in 5 mL cultures of antibiotic media for plasmid extraction and verification by PCR. PCR analysis confirmed that plasmids previously used for transformation did contain the expected insert of NDX8 (Fig, 4.2C), and therefore there likely was a problem with Colony PCR. Bl21 (DE3) cells are known to grow much faster than typical subcloning cells such as DH5α or, in this case, High-Control 10G cells, and it's possible that large colonies observed after transformation may have had some effect on the Colony PCR protocol itself.

Assuming that majority of the colonies obtained from transformation of pET-NDX8 into BL21 (DE3) were positive based on their grown in the presence of kanamycin and despite the negative Colony PCR results, three cultures were grown 5 mL antibiotic media for small scale expression analysis, grown at 37 °C until OD_{600} was ~0.5, whereupon expression was induced with the addition of 1 mM IPTG and cultures were incubated overnight shaking at 16 °C. Cells were harvested, resuspended in 250 µL NBB, and lysed at 95 °C for 10 minutes. Cell lysates were pelleted, supernatant was separated from pellet fraction, and pellet was resuspended in 20 µL 4X Laemmli buffer. Supernatant and pellet fractions were analyzed by 12% SDS-PAGE as soluble and insoluble protein fractions,

and a band in the pellet fraction was identified near the 25 kD marker to indicate expression of the 27.2 kD NDX8 protein (Fig. 4.2D).

The insoluble expression of NDX8 was unable to be remedied by the assessment of different cultures, different incubation temperatures and times, or different concentrations of IPTG as an inducer. Considering that the activity of the enzyme is dependent on its solubility, and the appropriation of large amounts of active enzyme are dependent on soluble expression, the expression of recombinant NDX8 would need to be carried out with a fusion partner to aid its solubility.

4.2.1.2 Cloning and Expression of SUMO-NDX8

The addition of a SUMO fusion to NDX8 to increase its solubility during expression was the considered the path of least resistance since a similar recombination cloning system was provided by Lucigen with the SUMO fusion encoded upstream of the recombination site in the linearized pETite-SUMO-nHis-Kan vector. PCR was carried out on the NDX8 geneblock with primers to add 18 bp flanking regions homologous to the SUMO-containing pETite vector (Fig. 4.3, 4.4A), and transformation was carried out similarly as before to produce the pETS-NDX8 plasmid confirmed in 100% of screened colonies by Colony PCR (Fig. 4.4B). PCR analysis of extracted plasmids confirmed insertion of the NDX8 gene after the SUMO fusion (Fig. 4.4C).

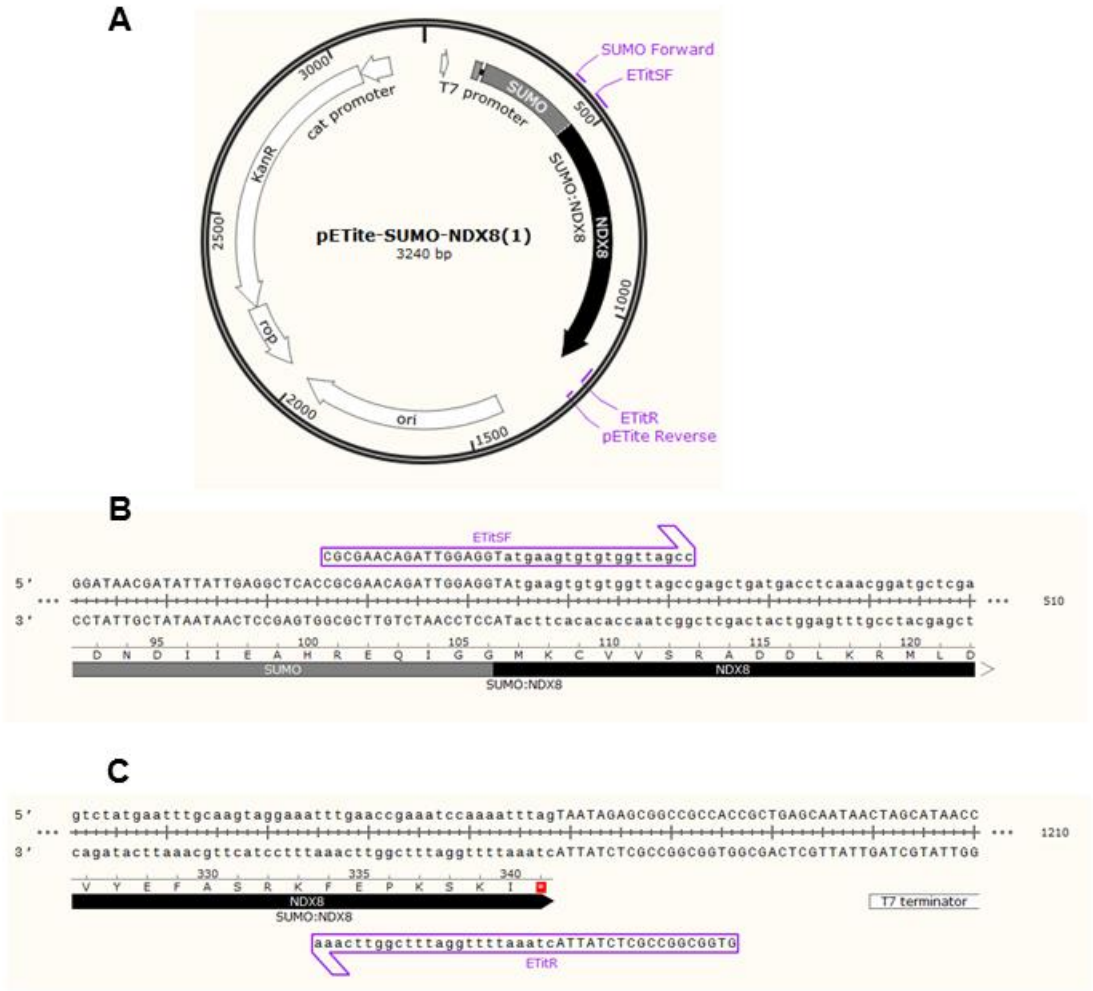


Figure 4.3 Plasmid design of pETite-SUMO-NDX8

The primers designed for the generation of the pETite-SUMO-NDX8 plasmid (A) were intended to overlap ~18 bp of the NDX8 gene along with the vector on each end of the insert (B, C). The inclusion of the start and stop codons as well as the vector sequence in the plasmid insures in-frame insertion of the gene of interest through the ligation-independent recombination method.

Transformation of the PCR-confirmed pETite-SUMO-NDX8 plasmid into High-Control BL21 (DE3) cells successfully produced transformants on solid antibiotic media, and small scale expression was carried out using 1 mM IPTG

added at OD_{600} ~0.5 to confirm the successful expression of SUMO-NDX8 by SDS-PAGE (Fig. 4.5A). When cells were lysed similarly as before, the SUMO-NDX8 band of 38.8 kD showed up in the pellet fraction by SDS-PAGE analysis. Considering that heating is a source of denaturation, another small scale expression was carried out where cell lysis was achieved via sonication with a probe-tip sonicator while the sample was in ice. Similar to the heat lysis protocol, the SUMO-NDX8 protein was found in the pellet fraction upon SDS-PAGE analysis (Fig. 4.5B). More small scale expressions were carried out using sonication for lysis and varying the concentration of IPTG from 5 μM to 500 μM, however the results remained that SUMO-NDX8 was insoluble (Fig. 4.5C).

The insolubility of SUMO-NDX8 was unable to be altered by experimental parameters such as the changing of inducer concentration, incubation temperature, or time of expression. The possibility of purifying the protein in a denatured protocol was an option, but the subsequent refolding of the protein would be a potential barrier that could be insurmountable considering that the protein itself is an enzyme that contains multiple cysteine residues. Therefore regaining the native and active structure of the enzyme was something that could be a problem. Based the existence of these cysteines, the addition of a thioredoxin fusion protein was considered.

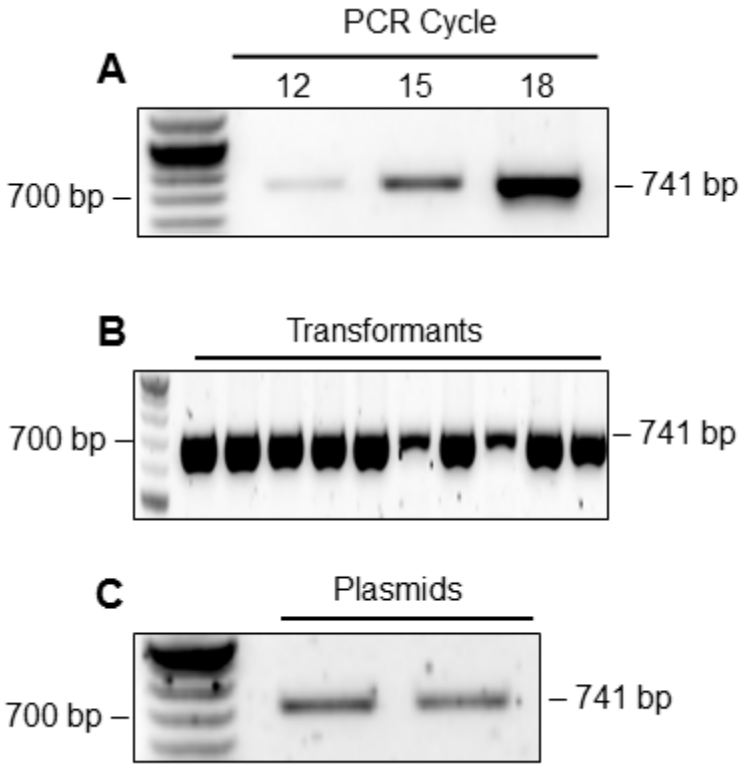


Figure 4.4 Cloning of SUMO-NDX8

A, PCR amplification of NDX8 geneblock to generate expected 741 bp fragment with 18 bp overlaps with linear pETite-nHis-SUMO vector. B, Colony PCR on High-Control 10G cells using cloning primers to amplify the 741 bp insert. C, PCR on extracted plasmids from High-Control 10G cells with specific primers that verify insertion downstream of SUMO fusion

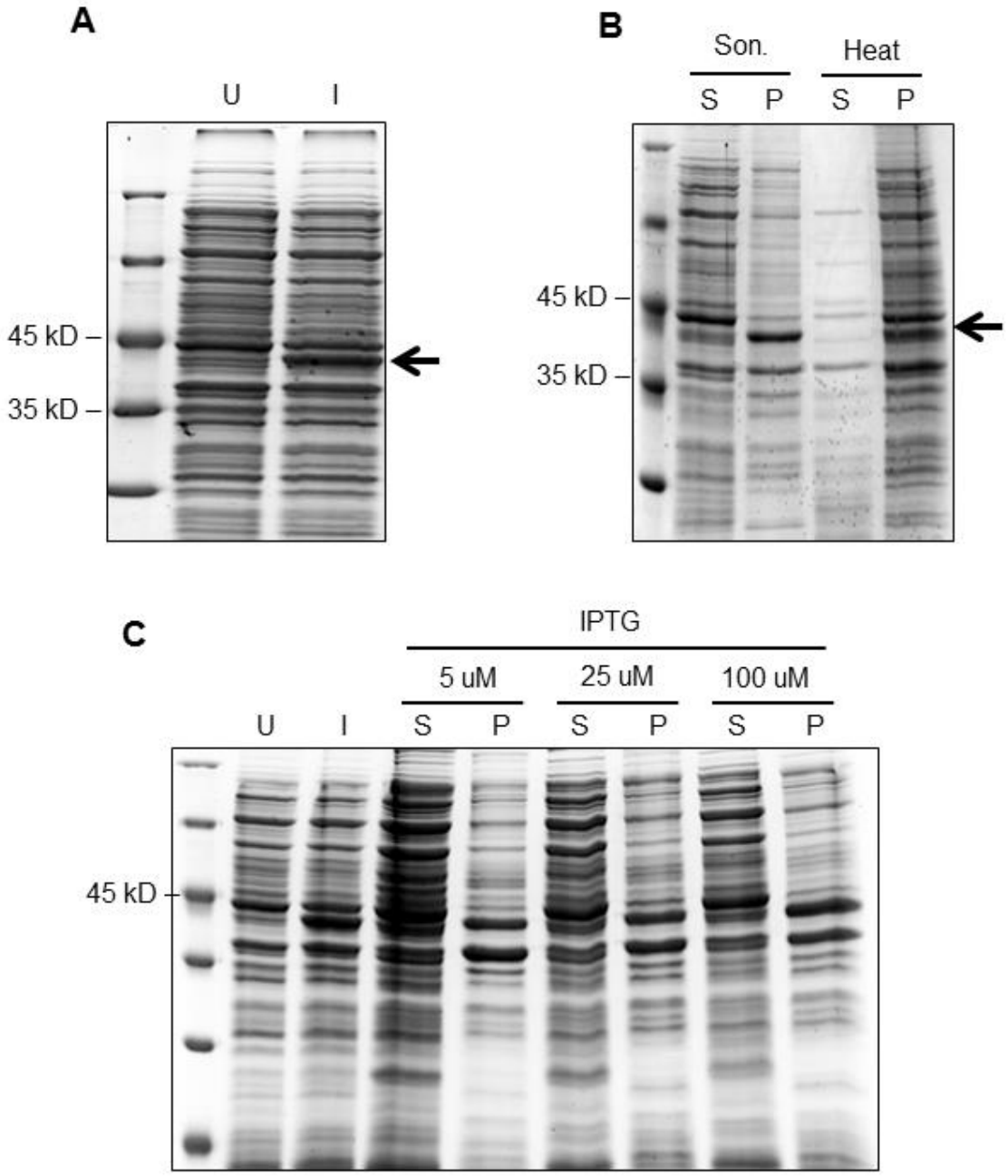


Figure 4.5 SUMO-NDX8 Protein Expression

A, SDS-PAGE of small scale protein expression showing induced expression of the SUMO-NDX8 protein approximately 39 kD in size. B, SDS-PAGE of solubility analysis comparing sonication and heating as two separate methods for cell lysis, with soluble (S) and pellet (P) fractions indicating protein is insoluble. C, SDS-PAGE of small scale expression using lower concentrations of inducer IPTG to enhance solubility showing soluble (S) and pellet (P) samples after sonication to lyse cells.

4.2.1.3 Cloning and Expression of TrxA-NDX8

The previous investigators who first characterized the *C. elegans* NDX8 protein did so using recombinant protein expressed in *E. coli* with a thioredoxin (TrxA) fusion (AbdelRaheim and McLennan 2002). Although SUMO has been regarded as a superior fusion partner to enhance solubility of expressed recombinant proteins, each protein behaves differently and thus responds differently to alternate fusion partners (Malakhov et al. 2004). The thioredoxin fusion protein, on the other hand, has been stated to aid in the formation of cysteine disulfide bonds and additionally increase the solubility of recombinant proteins (LaVallie et al. 1993). Therefore pET-32b was obtained from the Rangachari lab that contained an in-frame TrxA fusion and was modified for our purposes of inserting NDX8.

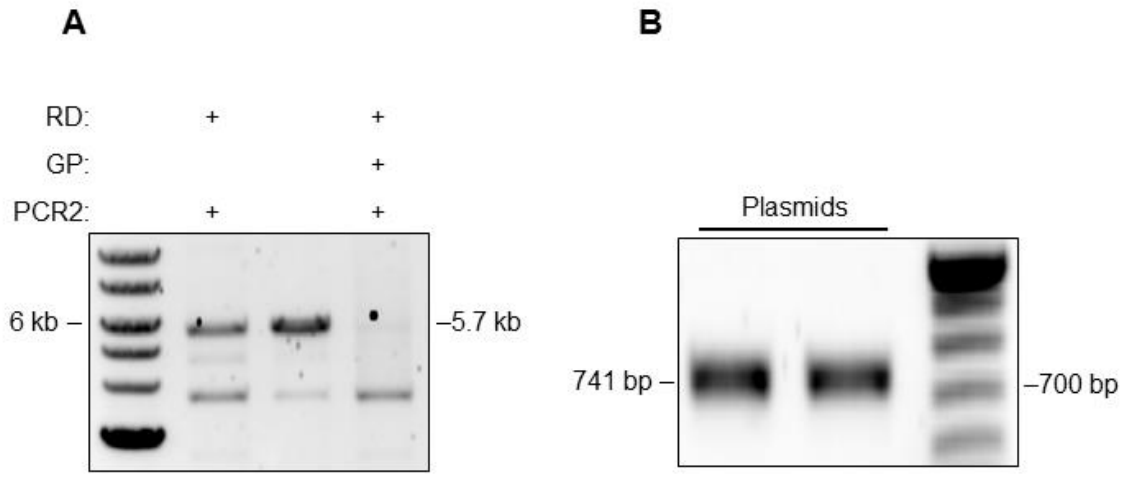


Figure 4.6 Cloning of TrxA-NDX8

A, PCR amplification of pET-32b backbone with restriction digestion (RD), gel purification (GP), or secondary PCR (PCR2). B, PCR amplification on plasmids extracted from High-Control 10G cells with specific primers to verify the insertion of NDX8 downstream of TrxA fusion.

The pET-32b plasmid gifted by the Rangachari lab contained an insert for their research interests. The backbone therefore needed to be modified to remove that insert for the cloning of NDX8 to be successful. The linearized empty backbone of the pET-32b plasmid generated by PCR using primers designed to amplify away from the insertion of the previous gene (Fig. 4.1, 4.6A). The PCR amplification will increase the amount of linearized vector orders of magnitude above the original template, but this template can still interfere with transformation downstream and cause a heterogeneous transformant population to lower the efficiency of the cloning. Although attempts were made to digest the plasmid within the previous insert before and after PCR with the restriction enzyme *SacII*, PCR on the given and undigested pET-32b showed the best results in terms of intensity and homogeneity of target (Fig. 4.6A).

The NDX8 geneblock was PCR amplified to add flanking 18 bp regions of homology to the newly amplified backbone of pET-32b. The modified NDX8 fragment and linearized vector were combined in High-Control 10G cells for transformation, and the next day Colony PCR confirmed the production of the pET32-NDX8 plasmid in 95% of transformants. Plasmids were subsequently extracted from positive cultures and used as templates for PCR to confirm the insertion of the NDX8 gene downstream of the TrxA fusion protein (Fig. 4.6B).

The PCR-verified plasmids were used transform Rosetta 2 (DE3) cells for protein expression. The Rosetta cell line was used in lieu of Bl21 (DE3) cells for its chloramphenicol-selected plasmid that encodes rare tRNAs that aids in the expression in bacteria of eukaryotic proteins without codon optimization. Colony

PCR verified 100% of Rosetta 2 (DE3) transformants analyzed contained the plasmid with the gene-of-interest, and subsequent small scale expression using 0.5 mM IPTG showed positive expression of the 41 kD TrxA-NDX8 fusion by SDS-PAGE (Fig. 4.7A), however subsequent cell lysate analysis by SDS-PAGE showed that the fusion was, yet again, insoluble.

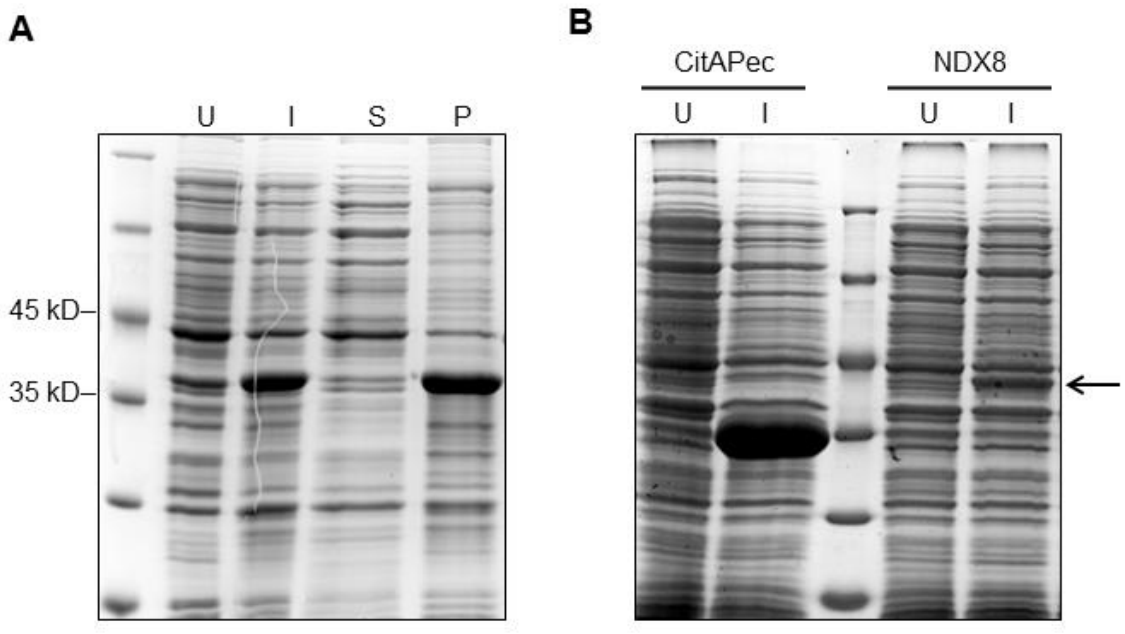


Figure 4.7 TrxA-NDX8 Protein Expression

A, SDS-PAGE of solubility assay on small scale expression of 41 kD TrxA-NDX8 from Rosetta 2 (DE3) cells. B, Comparison of proteins CitAPec and NDX8 expressed from the same vector in a T7 system to illustrate the low yield of NDX8 in crude cell lysate.

The inability to produce *C. elegans* NDX8 in a soluble form in high yield was a perplexing road block, especially considering that a soluble protein had been obtained previously (AbdelRaheim and McLennan 2002). Despite altering concentrations of inducer, induction temperatures, and fusion partners, the NDX8

protein remained insoluble during preparation. In addition to the insolubility of the protein, the yields observed in crude extracts were significantly lower in comparison to other previously expressed proteins using the same T7 promoter (Fig. 4.7B). It is unknown whether this difference in expression was due to the protein activity itself or some other factor. NDX8 was previously described as the most efficient Nudix CoAse characterized to date, with a K_{cat} of 13.8 s^{-1} and a K_{cat}/K_m of 64 mM^{-1} s^{-1} (AbdelRaheim and McLennan 2002). Because CoA is such a vital component of cellular metabolism, the possibility exists that the activity of this highly efficient recombinant enzyme in a soluble form was detrimental to the cell growth, and only cultures where the protein was expressed insolubly or sequestered into an inclusion body therefore survived to be harvested for protein extraction. It is peculiar, however, that despite analysis showing the protein was insoluble regardless of which expression parameter was varied, the total recombinant protein generated was quite low considering the use of the T7 promoter. Assuming that transcription from the T7 promoter proceeded at a comparable rate to other recombinant proteins using the same system, the purification of soluble NDX8 protein may result from some proteolytic event (Baneyx and Mujacic 2004). Regardless of the mechanism responsible, the solubility issues of recombinant NDX8 expression caused the protein to be deemed not to be a viable option for accumulating a Nudix CoAse for use in a specific CoA-RNA capture protocol.

4.2.2 M. musculus Nudt7

4.2.2.1 Cloning and Expression of SUMO-Nudt7

The Nudix CoAse Nudt7 from *M. musculus* was characterized to be less efficient than *C. elegans* NDX8 with a K_{cat}/K_m of 16 mM^{-1} s^{-1} (Gasmi and McLennan 2001). The differences in activity from NDX8, however, may be the key to its purification as well as its use in the CoA-RNA capture protocol. The lower enzyme efficiency of Nudt7 may allow the survival of cells solubly expressing the active enzyme. In addition, the activity of Nudt7 against CoA-thioesters implies the distinct possibility that it may be better suited to release CoA-RNA from our CoA-RNA capture protocol (Reilly et al. 2008)

The Nudt7 gene was obtained from a mouse embryonic fibroblast (mEF) cDNA library from the Guo lab that was library generated by random hexamer primers. The cDNA library is an advantageous source of cloning eukaryotic proteins because of its generation from captured mRNA. These mRNAs have already undergone splicing and serve as true translation substrates. PCR was used to amplify the Nudt7 transcript from the cDNA library using primers designed to add flanking 18 bp sequences homologous to the pETite-SUMO-nHis-Kan vector (Fig. 4.8A), and the resulting fragment was added directly with linearized vector for recombination and transformation into High-Control 10G cells. Colony PCR with general primers to amplify any insert present confirmed the correct incorporation into the pETite-SUMO vector in 70% of screened transformants (Fig. 4.8B), which was similarly confirmed by the subsequent PCR on extracted plasmids using gene-specific primers (Fig. 4.8C).

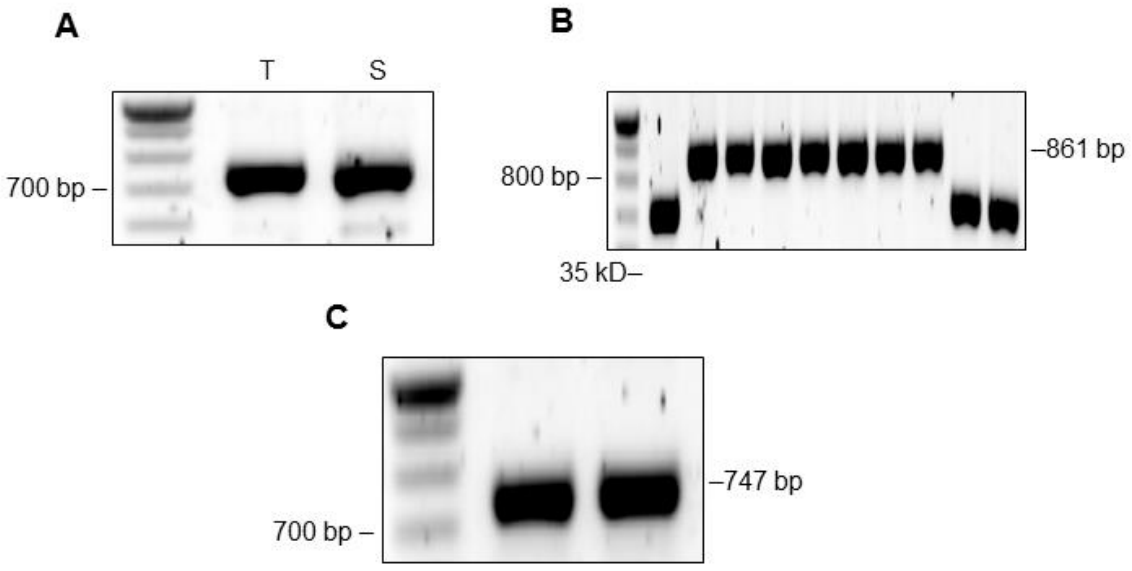


Figure 4.8 Cloning of SUMO-Nudt7

A, PCR amplification to prepare Nudt7 for cloning with TrxA fusion (T) and SUMO fusion (S). B, Colony PCR of High-Control 10G cells with general backbone primers to confirm insertion of Nudt7 gene into pETite-SUMO vector. C, PCR analysis of extracted plasmids from High-Control 10G cells with specific primers to confirm insertion of Nudt7 downstream of SUMO fusion.

The newly generated pETS-Nudt7 plasmid was transformed into High-Control Bl21 (DE3) cells, whereupon Colony PCR analysis using gene-specific primers showed 60% of transformants screened contained the correct gene. Small scale expression of Nudt7 followed by solubility assay showed that the protein was expressed at a higher rate than NDX8 previously was, and although majority of the protein was present in the insoluble fraction, there was a noticeable band below the 45 kD marker that indicated some of the 39 kD SUMO-Nudt7 was solubly expressed (Fig. 4.9A). Large scale expression was carried out in 1 L of antibiotic media where expression was induced at an OD_{600} of 0.4 by addition of 0.5 mM IPTG and carried out overnight at 16 °C. Upon SDS-

PAGE analysis of protein extraction, the majority of the protein was observed to be soluble (Fig. 4.9B). Purification of the N-terminal His_6-tagged protein by IMAC resin was carried out to apparent homogeneity using 250 mM imidazole as the eluant. Samples were combined and dialyzed against 1X PDB overnight at 4 °C to remove imidazole and salt, and afterwards quantity was determined by Beer's law using absorbance at 280 and the molar extinction coefficient calculated to be 20,650 M^{-1} cm^{-1}, resulting in a total of 12 mg L^{-1} protein purified at a concentration of 160 μM.

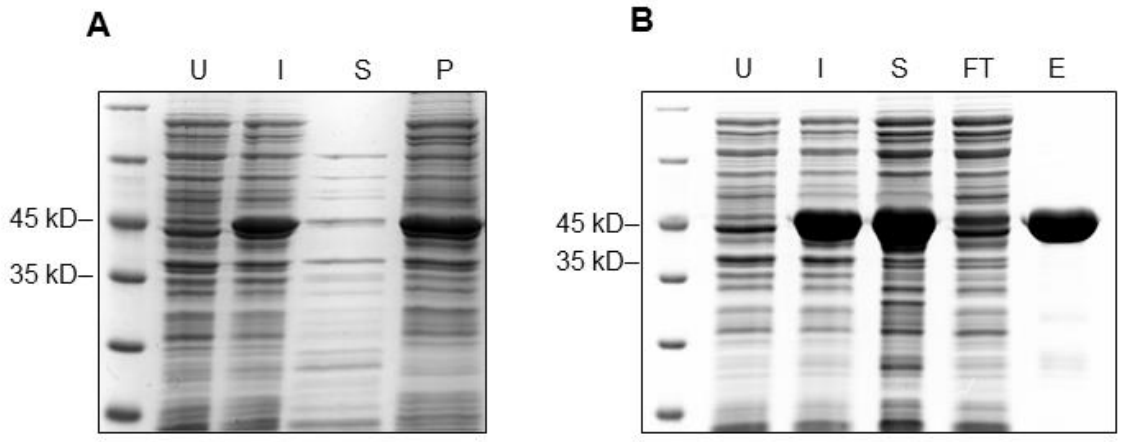


Figure 4.9 SUMO-Nudt7 Protein Expression

A, SDS-PAGE of small scale protein expression with uninduced (U) control, induced (I) lysate, soluble (S) extract, and insoluble (P) extract showing 39 kD SUMO-Nudt7 to be insoluble. B, Purification of large scale SUMO-Nudt7 expression showing uninduced control (U), induced cell lysate (I), soluble cell lysate (S), affinity column flow-through (FT), and final eluted protein (E).

4.2.2.2 Nudt7 Activity

The activity of SUMO-Nudt7 against 1 mM CoA was carried out in 1X Nudix buffer at pH 7.5 for 1 h at 37 °C. The reaction was visualized by HPLC

using a strong anion exchange column with an isocratic elution protocol to separate products from reactants based on the different number of phosphate groups of each while monitoring the absorbance at 260 nm of the adenosine group (Fig. 4.10A). CoA hydrolysis was observed as the appearance of a distinct peak belonging to the product 3’5’-ADP appearing with a lower retention time than CoA itself (Fig. 4.10B). The resolution of the two peaks varies with the percentage of 0.4 M KH_2PO_4 used, but the isocratic elution at 40% phosphate resulted in a general 2.5 minute difference in retention times between CoA and 3’5’-ADP.

To test the activity of Nudt7 against CoA-RNA, *In Vitro* Transcription with α-^{32}P-ATP was carried out in the presence of depCoA to generate a 5’ CoA-RNA of 120 nt length following previously a described protocol (Huang 2003a). The subsequent RNA was purified from buffer salts with an M30 column and used for biotinylation in the presence of imidazole following a previously described protocol (Coleman et al. 2005). The resulting Biotin-CoA-RNA was incubated for 5 minutes at room temperature with streptavidin and resolved by 8% denaturing PAGE for analysis. Nudt7 digestion of RNA was carried out both before and after biotinylation in separate reactions, and the successful hydrolysis was observed as a lack of electrophoretic mobility shift on the gel (Fig. 4.10C).

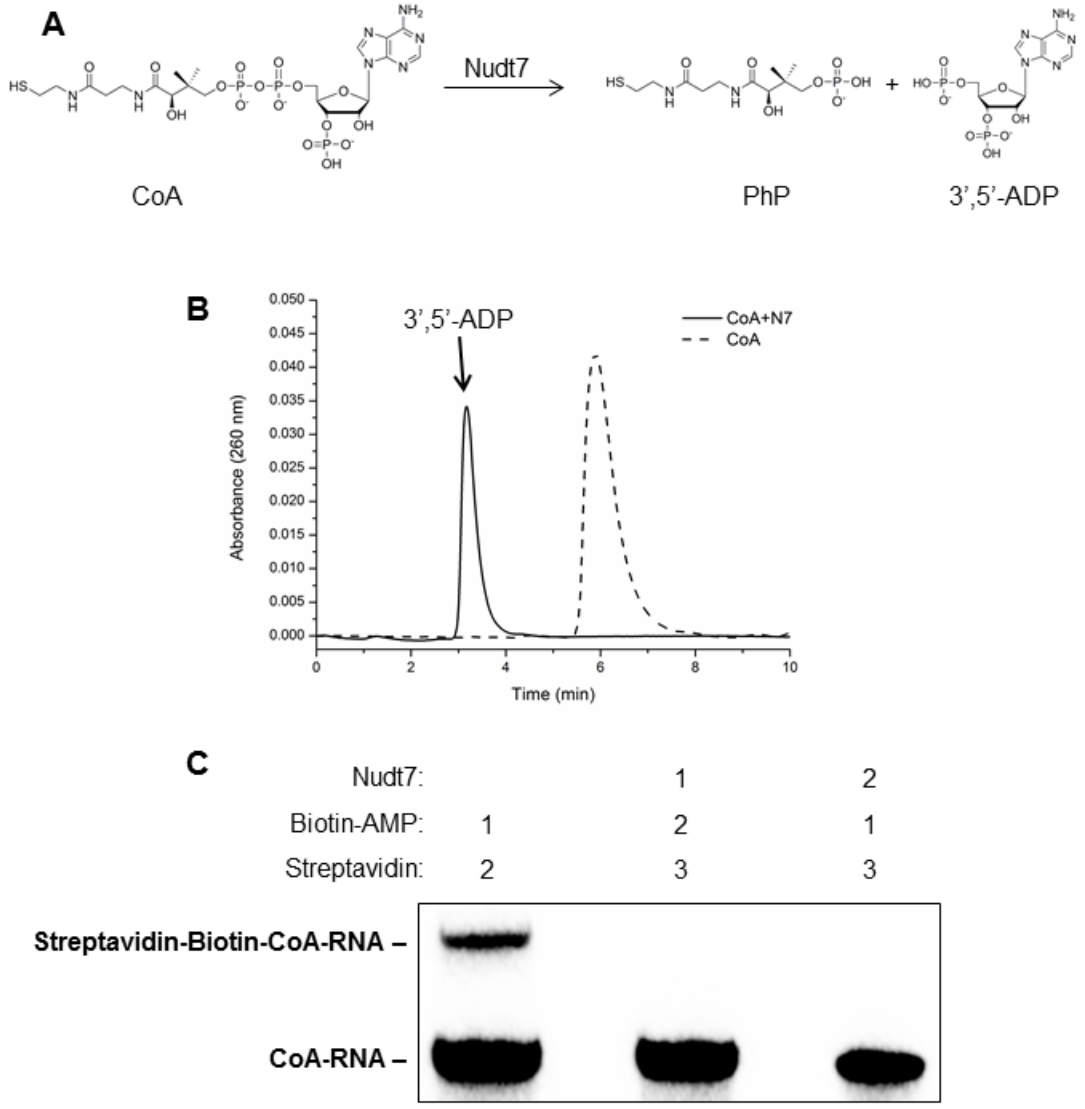


Figure 4.10 Nudt7 Activity

A, Scheme of Nudt7 hydrolysis of CoA to produce 4'-phosphopantetheine (PhP) and 3', 5'-adenosine diphosphate (3', 5'-ADP). B, Chromatogram of Nudt7 reaction with CoA monitored by HPLC using isocratic elution on strong anion exchange column. C, Autoradiograph of Nudt7 activity against internally labeled CoA-RNA before and after biotinylation with the number indicates the order of treatments.

Interestingly, Nudt7 was capable of hydrolyzing all CoA after biotinylation. The ability for Nudt7 to hydrolyze biotinylated CoA may be synonymous to the reported increasing efficiency for cleaving CoA thioesters up to C_{12} in length

where the K_{cat}/K_m peaks at 22.26 mM^{-1} s^{-1}, an approximately 40-fold higher efficiency than for CoA itself (Reilly et al. 2008). This trait would make Nudt7 a great candidate for the specific release of CoA-RNA in a capture protocol. While most conceivable CoA-RNA capturing schemes would involve specific thiol chemistry such as the formation of a covalent disulfide or a thioether bond and would exclude the capture of other coenzymes, namely NAD^+, the activity of Nudt7 against NAD^+ and FAD was assessed to confirm its substrate range to be specifically CoA and it's thioesters. Nudt7 was incubated at a concentration of 16 µM with 1 mM NAD^+ and 1 mM FAD for 1 h at 37 °C in 1X Nudix buffer. Reactions were analyzed by HPLC using isocratic elution from strong anion exchange resin to separate peaks based on the difference in phosphates contained by products and reactants, and as a result Nudt7 was determined to have no activity against FAD or NAD (Fig. 4.11C and D).

.

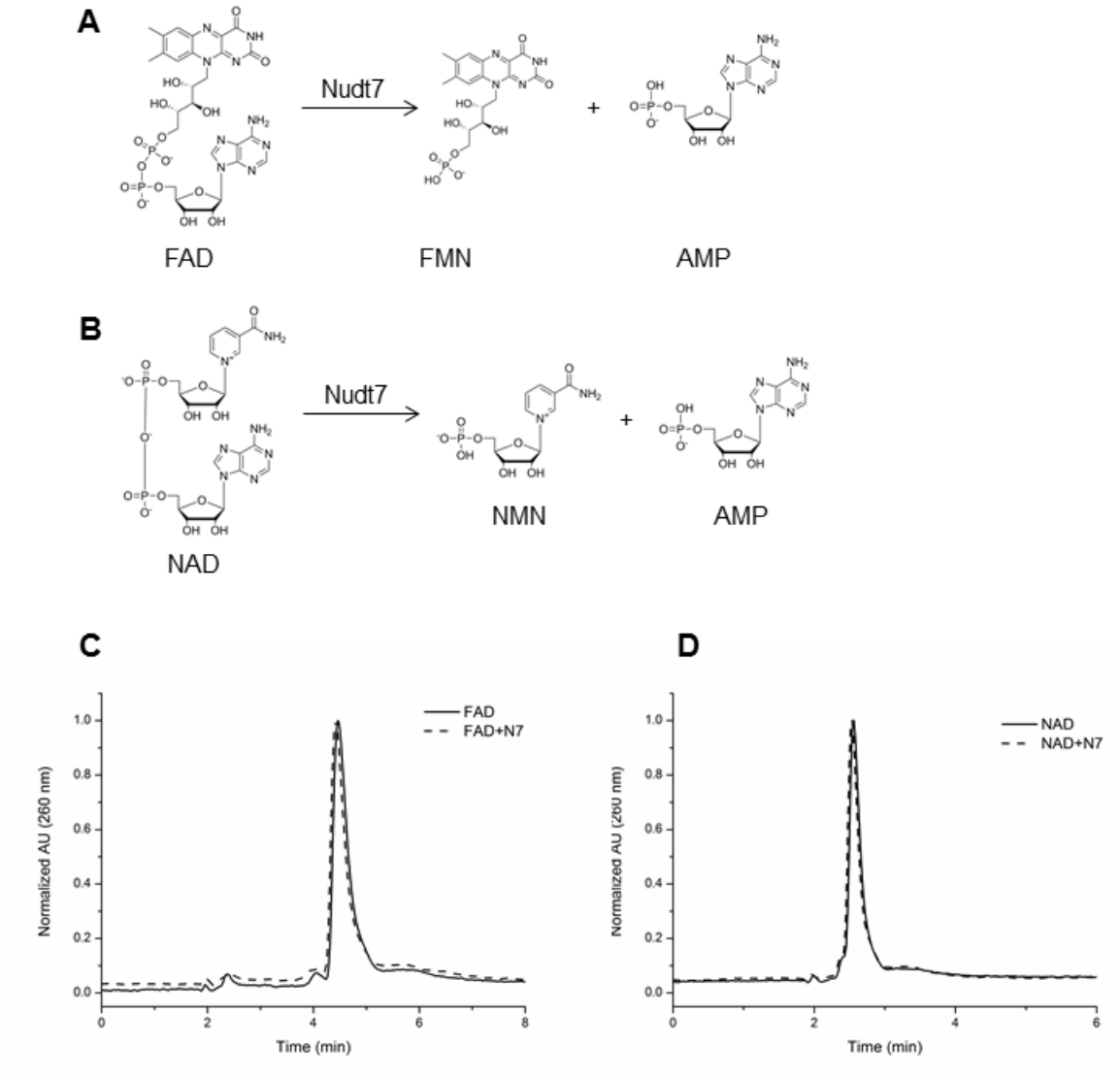


Figure 4.11 Nudt7 Activity Against NAD and FAD

A, Scheme of proposed Nudt7 activity against FAD to produce Flavin mononucleotide (FMN) and adenosine monophosphate (AMP). B, Scheme of proposed Nudt7 activity against NAD to produce nicotinamide mononucleotide (NMN) and adenosine monophosphate (AMP). C, Chromatogram of Nudt7 reaction with FAD showing no difference in retention time using isocratic elution on strong anion exchange column. D, Chromatogram of Nudt7 reaction with NAD showing no difference in retention time using isocratic elution on strong anion exchange column.

Based on the protein solubility, quantity purified, and activity demonstrated, Nudt7 was considered to be a viable candidate for use in specific CoA-RNA capture protocols without the removal of the SUMO fusion. The high activity towards CoA and thioesters such as our *in vitro* made biotin-CoA indicate that CoA-RNA isolated from bacterial cells can be modified with long linkers via thioester or thioether formation, where the free terminus contains a biotin moiety capable of capturing on streptavidin resin. Nudt7 can thereby be used to specifically release such CoA-RNA from the resin with a 5'-phosphate ready for ligation and subsequent reverse transcription. The inactivity of Nudt7 against NAD and FAD insure that, in the highly unlikely event where other coenzyme-RNA conjugates are immobilized on a resin in the capture protocol, only CoA-RNA will be released by Nudt7. Considering that 1 h in the initial reactions was more than enough time to hydrolyze 1 mM CoA, the Nudt7 release step of the CoA-RNA capture protocol can likely be completed in shorter amounts of time, especially when considering isolated amounts of CoA-RNA are estimated to be far below 1 mM (Kowtoniuk et al. 2009).

4.2.3 E. coli NudC

4.2.3.1 Cloning and Expression of MBP-NudC

Investigators have previously shown that NudC from *E. coli* is capable of hydrolyzing CoA in addition to both oxidized and reduced NAD (Bird et al. 2016). We therefore reasoned that if this enzyme is capable of utilizing as substrates two structurally unrelated coenzymes such as these, then it likely also can utilize FAD as a substrate. Investigation of the enzyme kinetics of NudC with these

various substrates may further indicate its role as a general coenzyme hydrolase within *E. coli* that has a slight catalytic preference for NAD.

NudC was prepared for cloning by PCR of the *E. coli* genome obtained from the discarded fraction of a mini-prep kit carried out on DH5α cells. Mini-prep kits are fairly standard in lab protocols, where the initial steps include resuspension of cells in a buffer, addition of a lysis buffer typically containing SDS and NaOH, and the subsequent precipitation of SDS by the addition of a buffer containing chaotropic salts such as guanidinium HCl. The SDS is used to solubilize the lipid cell membrane to aid lysis, which incidentally is closely associated with the bacterial genome. As the purpose of mini-prep kits is to isolate cytosolic plasmid DNA, the lipid portion is selected against by precipitating the SDS-solubilized lipids by the addition of the chaotropic agent. Following the precipitation, the suspension is clarified by centrifugation, the soluble fraction containing the cytosolic plasmid DNA is taken further for purification, and the pellet fraction containing the bacterial genome is typically discarded. Here, we took a portion of the precipitated lipid and genome fraction and resuspended diluted it in water for use as a PCR template.

The recovered *E. coli* genome was used for PCR with primers to specifically amplify NudC and add flanking 18 bp regions that are homologous to the pMBP-Parallel vector in frame with the fusion protein (Fig. 4.12B). In addition, pMBP-Parallel was prepared by PCR amplification of the backbone with specific primers that were designed linearize the vector so that it could be used for the insertion of any gene downstream of the MBP fusion (Fig. 4.12A). The successful

recombination of the PCR amplified NudC from the *E. coli* genome and the modified pMBP-Parallel backbone would result in a plasmid construct capable of producing the maltose-binding protein fusion, followed by a thrombin cleavage site and the inserted NudC gene. The PCR-prepared backbone and insert were combined in DH5α cells for transformation, and subsequent Colony PCR analysis showed a strikingly low amount of positively recombined transformants. One single colony out of 33 tested showed evidence of successful recombination (Fig. 4.12C). Upon further inspection of the extracted plasmid by PCR amplification using general vector primers, the recombination to generate the plasmid pEMC was determined to be successful (Fig. 4.12D).

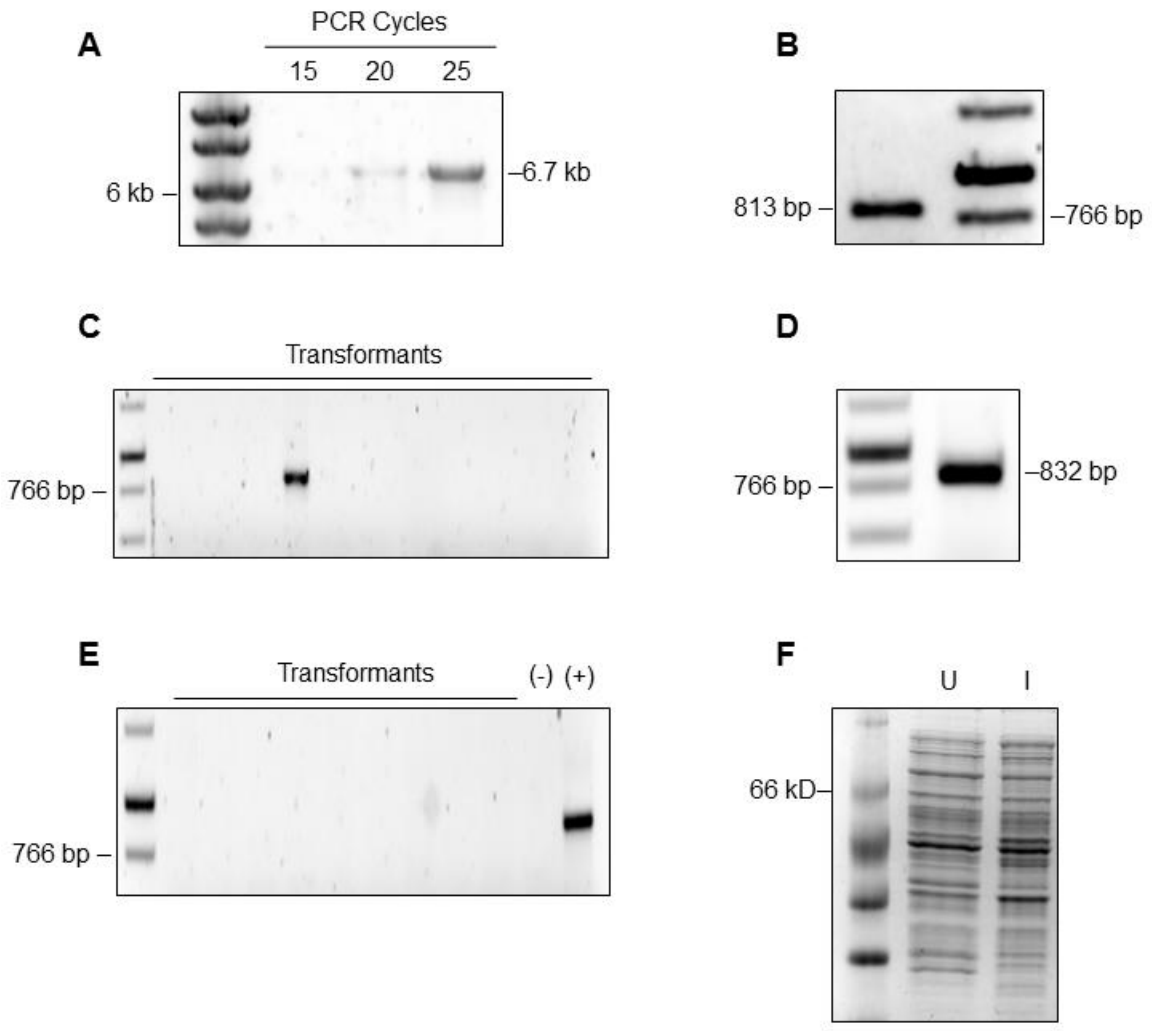


Figure 4.12 Cloning of MBP-NudC

A, PCR amplification at different cycle numbers of the pMBP-Parallel backbone in preparation for cloning. B, Amplification of the NudC gene from the *E. coli* genome with 30 cycles of PCR and specific primers to add 18 bp overlap regions to pMBP-Parallel for recombination during transformation. C, Colony PCR of DH5α transformants using general primers for pMBP-Parallel. D, Verification of extracted plasmids from DH5α transformants using PCR with general primers for pMBP-Parallel. E, Colony PCR of NEBExpress transformants using general pMBP-Parallel primers, where PCR was carried out in batch along with water (-) and the pEMC plasmid used for transformation (+). F, SDS-PAGE of small scale expression from NEBExpress transformants where the 73 kD MBP-NudC was determined not to be expressed.

The subsequent transformation of pEMC into NEBExpress cells proved incapable of verification by Colony PCR (Fig. 4.12E). Considering the failure of Colony PCR had previously occurred with NDX8 constructs that upon further inspection were perfectly fine, small scale expression of MBP-NudC was carried out anyway only to result in a lack of observable protein expression (Fig. 4.12F). Despite repeating the procedure from each point individually, no expression of MBP-NudC was observed. Considering that PCR on the plasmid pEMC generated band at the expected size, the protein was never observed on SDS-PAGE. The primers used for the plasmid analysis are general primers that site outside of the insertion site and amplify whatever is between them. Typically in PCR, the appearance of a band at the expected size is verification of the gene being amplified, however this method cannot substitute for conventional sequencing techniques in the identification of mutations. It is therefore possible that the NudC gene was inserted properly into the vector, but some mutation arose that apparently inhibited protein expression. Such a situation could explain positive PCR results but negative expression of the protein.

4.2.3.2 Cloning and Expression of SUMO-NudC

The expression system of NudC was switched to the pETite vector considering the previous troubles encountered with the MBP fusion system. Based on the previous assumption that a mutation could have been responsible for the troubles with the previous MBP-NudC construct, the NudC gene was prepared again by PCR from the *E. coli* genome with specific primers to add flanking 18 bp regions homologous to vector pETite-SUMO-nHis-Kan (Fig.

4.13A). The amplified product combined with linearized vector in DH5α cells for transformation, and Colony PCR screening showed that 50% of analyzed transformants successfully had successfully undergone recombination to generate the pESC plasmid (Fig. 4.13B). Further analysis of extracted plasmids by PCR confirmed the insertion of the NudC gene downstream of SUMO encoded by the plasmid (Fig. 4.13C).

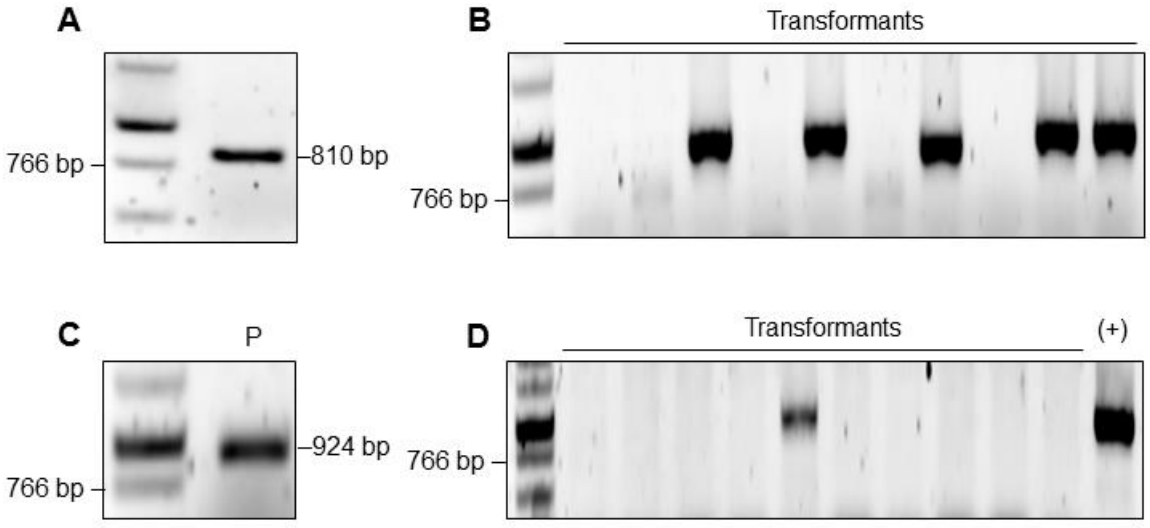


Figure 4.13 Cloning of SUMO-NudC.

A, PCR amplification of the NudC gene from the E. coli genome using primers that would add 18 bp regions of homology to the pETite-nHis-SUMO-Kan vector for recombination. B, Colony PCR analysis of DH5α transformants using general primers for the pETite-nHis-SUMO-Kan where a 924 bp fragment indicates incorporation of NudC downstream of the SUMO fusion. C, PCR on plasmids extracted from DH5α cells indicating that the pESC plasmid had been generated. D, Colony PCR analysis of High-Control Bl21 (DE3) transformants containing pESC with general backbone primers.

The verified pESC plasmid was used to transform High Control Bl21 (DE3) cells following a typical transformation protocol. Colony PCR analysis with the general vector primers showed that only 10% of screened transformants contained the gene-of-interest (Fig. 4.13D), and the one positive culture was

grown for small scale expression and solubility analysis. Expression under the T7 promoter was shown to be favorable, and the protein was observed in both the soluble and insoluble fractions of the cell lysate (Fig. 4.14A). When expressed at a larger scale, the protein remained soluble and was able to be purified by IMAC resin and eluted with 250 mM imidazole to apparent homogeneity (Fig. 4.14B). Quantity of protein was determined by Beer's law using the calculated extinction coefficient at 280 nm of 63,745 M^{-1} cm^{-1} and molecular weight of 41,971 Da, resulting in a total of 29 mg L^{-1} protein. The eluted samples were combined on an M30 column and washed with 1X PDB to reduce salt and imidazole concentrations before being stored at -20 °C in 30-40% glycerol

.

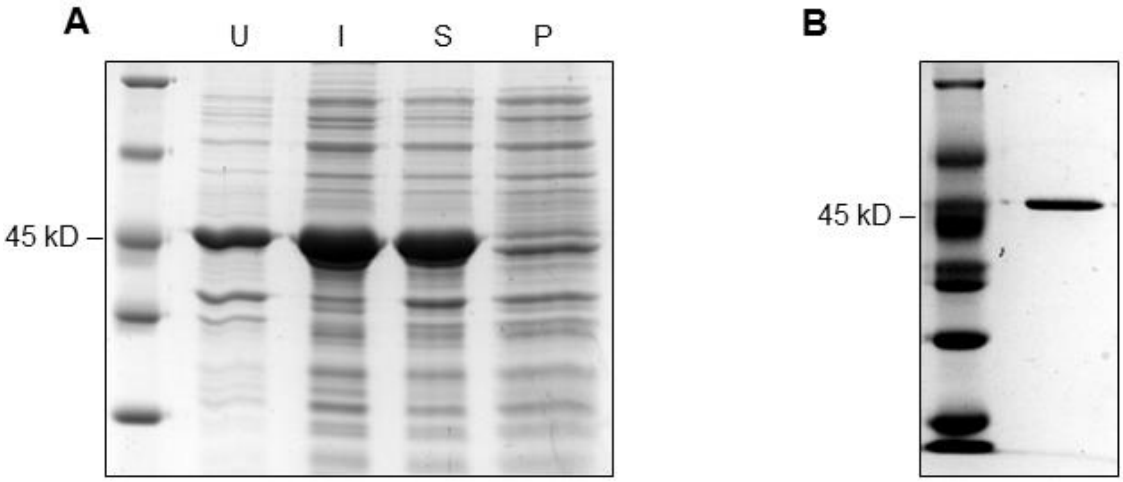


Figure 4.14 SUMO-NudC Protein Expression

. A, SDS-PAGE of solubility assay showing uninduced cell lysate (U), induced cell lysate (I), soluble cell lysate (S), and insoluble cell lysate (P) indicating that SUMO-NudC was expressed in the soluble fraction. B, SDS-PAGE of SUMO-NudC purified to apparent homogeneity.

4.2.3.3 NudC Activity

The activity of NudC was tested by incubation in 1X Nudix buffer with different coenzymes at 37 °C for 1 h. Analysis of the reactions by HPLC showed that 6 μM SUMO-NudC was capable of hydrolyzing 1 mM NAD^+ as indicated by previous investigators, but also 1 mM CoA and 1 mM FAD (Frick and Bessman 1995). Given the similarities in retention time between NAD^+ and its hydrolysis products of AMP and NMN, it is hard to quantify the rate the hydrolysis reaction (Fig. 4.15B). The general progress of the reaction can still be visualized by the splitting of the single NAD peak into its product peaks with low resolution. The progress of the FAD hydrolysis, on the other hand, is quite easy to follow by

HPLC because of the difference in retention times between FAD and the products FMN and AMP (Fig. 4.16B).

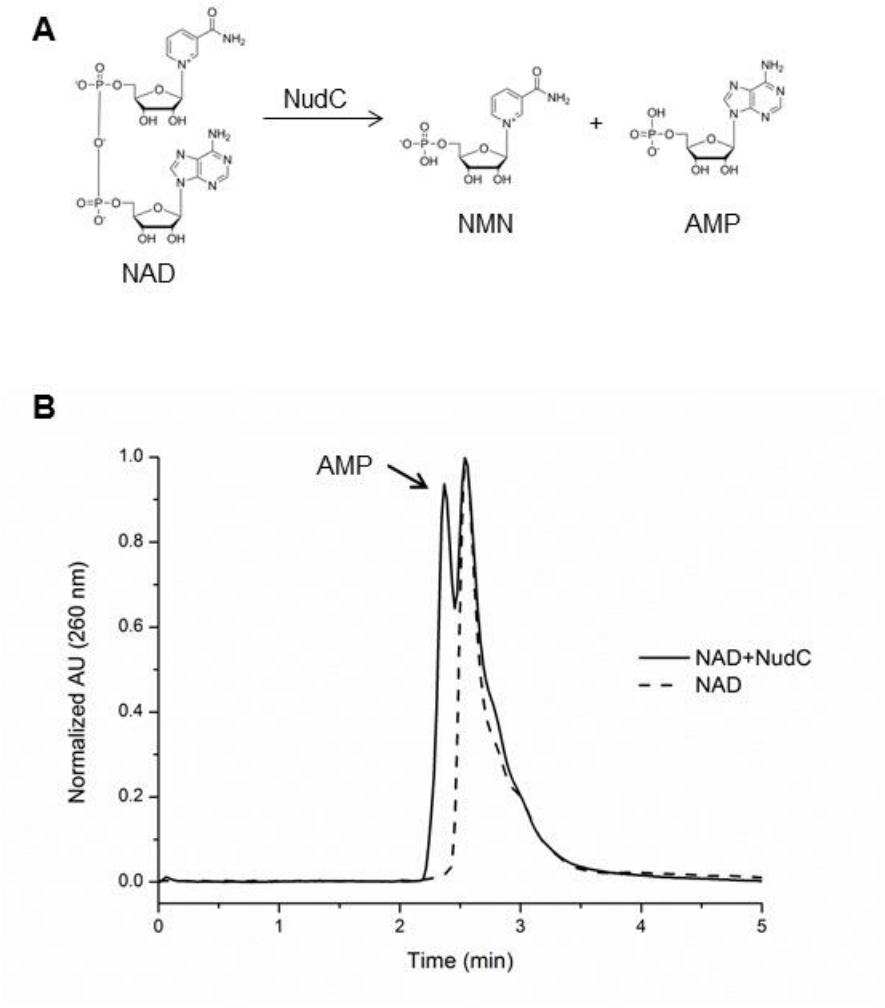


Figure 4.15 NudC Activity with NAD

A, Scheme of NudC activity against NAD to produce nicotinamide mononucleotide (NMN) and adenosine monophosphate (AMP). B, HPLC chromatogram showing appearance of AMP peak when NAD is treated with NudC, but is not well resolved from the original NAD peak.

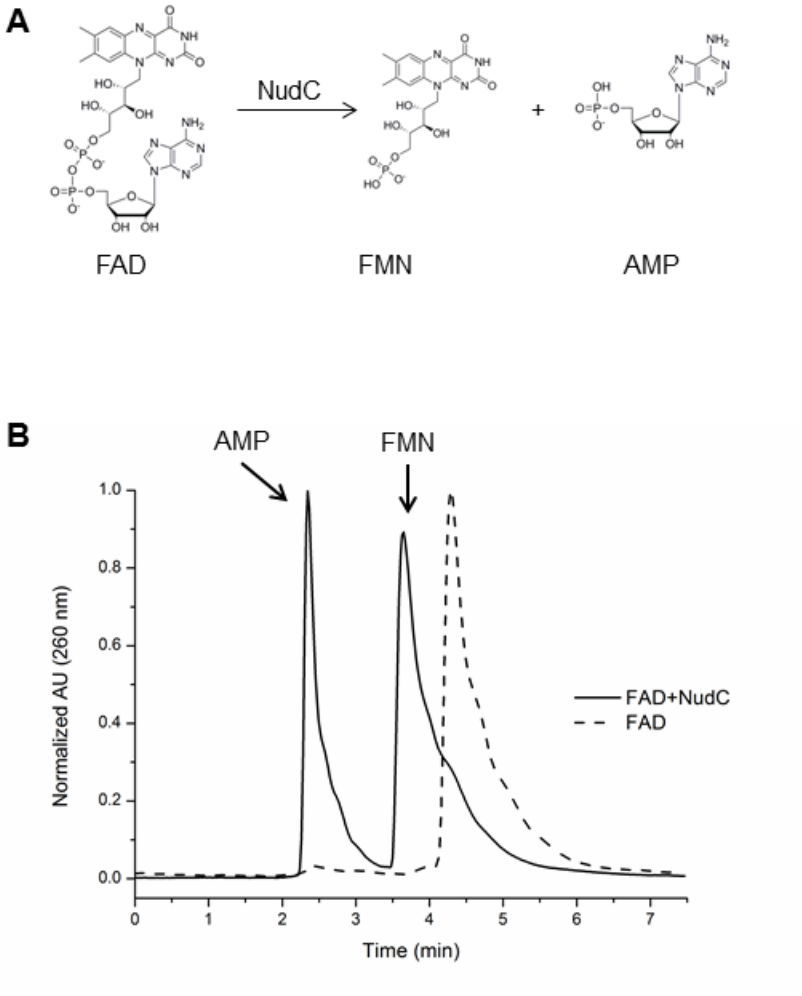


Figure 4.16 Figure 2 NudC Activity with FAD

A, Scheme of NudC hydrolysis of FAD into Flavin mononucleotide (FMN) and adenosine monophosphate (AMP). B, HPLC chromatogram showing reaction progress by the appearance of two separate peaks corresponding to AMP and FMN.

The previously uncharacterized NudC substrate of FAD is peculiar in that it seems, at first glance, to be hydrolyzed at a faster rate than NAD^+ (Table 4.1). As multiple reactions of NudC with FAD were carried out, it became apparent

that the K_m for FAD was below 0.05 μM, where incidentally the detection of all three peaks of FAD, FMN, and AMP becomes problematic in terms of quantifying rates of reaction using HPLC. The presumably lower K_m for FAD in addition to the previously reported K_m for NAD of 5.1 mM is likely the reason why the FAD is hydrolyzed at a faster rate than NAD at 1 mM each. The two coenzymes are not at equal concentrations in the cell where NAD is estimated around 1 mM and FAD near 100-fold lower, and therefore the enzyme may be tuned to accommodate either substrate equally in cellular conditions (Louie et al. 2003, Lin and Guarente 2003).

Table 4.1

Activity of NudC with FAD, NAD, and CoA

Substrate	**% Activity**	***K_m (mM)***
CoA	100	0.21
FAD	124	
NAD	5	5.1*
NADH		0.11*
AppA		0.67*
ADP-Ribose		1.8*

Activity calculated as percentage of rate (mM/min) at 15% reaction completion as determined by HPLC. *Denotes values obtained from Frick et al. (1995).

Table 4.2

Kinetic Constants of Nudix Hydrolases for CoA

Enzyme	***K_m (mM)***	***K_{cat} (s^{-1})***	***K_{cat}/K_m (mM^{-1} s^{-1})***
NudC	0.210	0.206	0.984
NDX8*	0.22	13.8	64
Nudt7**	0.240	3.8	16

Values for NudC were calculated with standard error under 0.1 *Data taken from AbdelRaheim et al. (2002). **Data taken from Gasmi et al. (2001)

The digestion of CoA by NudC was tested in a similar manner as Nudt7. Solutions of varying concentrations of CoA were incubated with NudC at 37 °C in 1X Nudix buffer, and reaction progress was monitored using HPLC (Fig. 4.17B). Kinetic studies revealed that NudC has a K_m for CoA of 0.210 mM, a K_{cat} of 0.206 s^{-1}, and therefore an enzyme efficiency of 0.984 mM^{-1} s^{-1} (Fig. 4.17C, Table 4.2). The K_m calculated for CoA falls between that previously calculated for NADH and NAD^+ (0.11 mM and 5 mM, respectively), and may reflect the enzymes compatibility with each substrate at intracellular concentrations (Frick and Bessman 1995, Lin and Guarente 2003, Miller et al. 2007). The half-saturation constant of NudC for CoA, however, is roughly equivalent to the reported value of murine Nudt7 (Gasmi and McLennan 2001). Despite the similarities in their half-saturation constants, the enzyme efficiency of NudC for CoA near 1 mM^{-1} s^{-1} is

16-fold lower than Nudt7 at 16.0 mM^{-1} s^{-1} in similar conditions, therefore indicating NudC to be slower enzyme than Nudt7 (Gasmi and McLennan 2001).

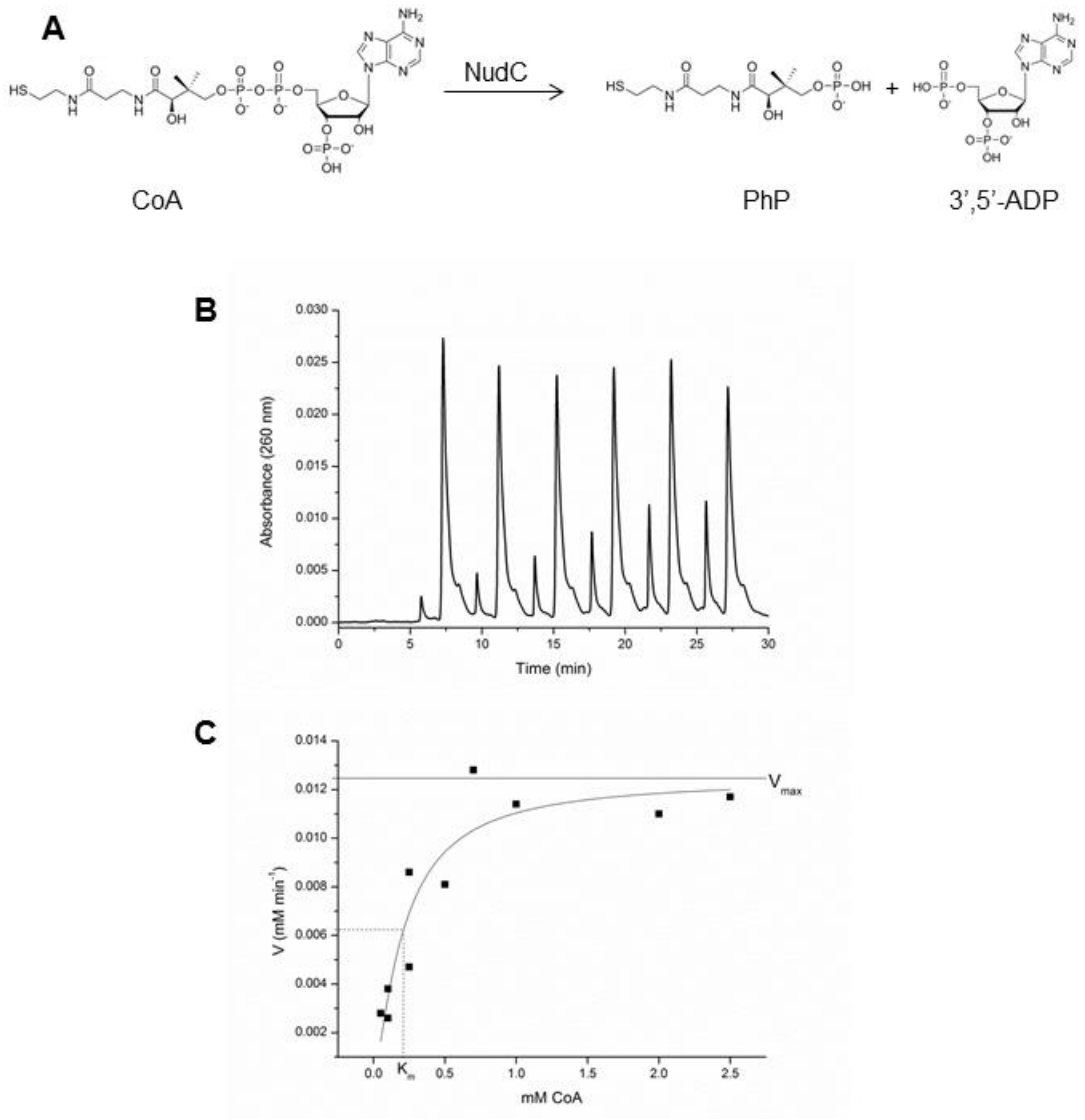


Figure 4.17 Activity and Kinetics of NudC with CoA.

A, Scheme of NudC hydrolysis of CoA to produce 4'-phosphopantetheine (PhP) and 3', 5'-adenosine diphosphate (3', 5'-ADP). B, Example of HPLC chromatogram used to monitor reaction progress of NudC and CoA. C, Michaelis-Menten plot of kinetic studies on NudC with CoA, with horizontal asymptote representing Vmax and dashed line representing the half-saturation constant Km.

The similarities in K_m between the Nudt7 and NudC are peculiar considering that NudC lacks the CoA consensus sequence upstream of its catalytic site that is found in all characterized Nudix CoA hydrolases. The lack of such a motif may explain the difference in efficiency between the two enzymes, indicating that the binding of CoA by NudC may be mediated predominantly by the adenosine portion of the molecule. Considering the activity of NudC against both NAD^+ and FAD (Fig. 4.15 and 4.16) as well as a host of other nucleotide derivatives, it's likely that NudC has a high affinity for adenosine itself, but displays compatible amino acids that specifically interact in a favorable manner with NADH that is reflected by a lower K_m (Frick and Bessman 1995). After all, a short 7 amino acid consensus sequence has been identified in Nudix NADH hydrolases that falls downstream of the catalytic site, and such a site may be responsible for the preferential binding of NADH over CoA (Xu et al. 2000). Upon its discovery, NudC was found to have a preference for NADH over NAD^+ showing about 5-fold higher activity for the reduced form (Frick and Bessman 1995). This difference in activity, as well as the 50-fold difference in K_m could reflect the ratio of free NADH/NAD^+ in the cell, and result from different interactions with the protein based on the charge of the nicotinamide group (Lin and Guarente 2003). Oxidized FAD, on the other hand, does not bear a positive charge, but displays its oxidized state by the presence of a pair of secondary imines in isoalloxazine ring (Fig. 4.18). This lack of a positive charge likely results in different protein-coenzyme interactions, resulting in an assumed lower K_m for

FAD than NAD^+ and leading to a higher activity of FAD hydrolysis when compared at equal concentrations.

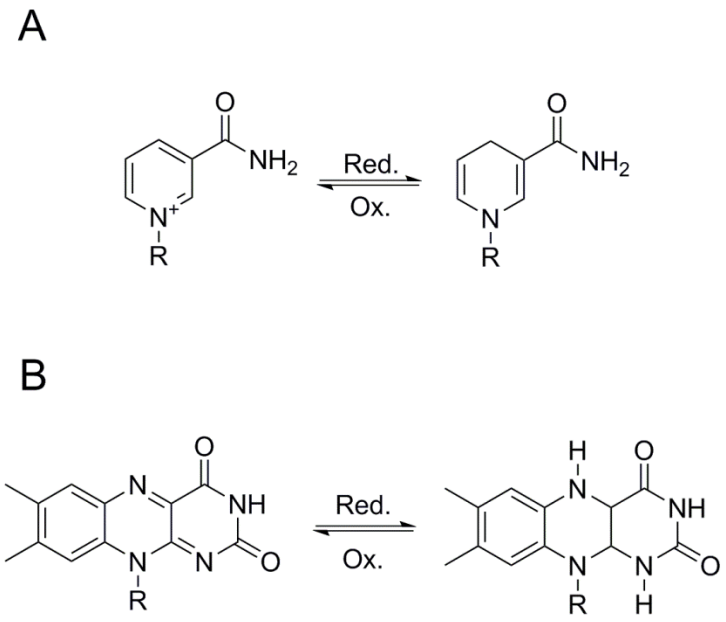


Figure 4.18 Redox States of NAD and FAD.

A, Redox states of NAD where the oxidized form has a negative charge on the nitrogen in the pyridine ring. B, Redox states of FAD where the oxidized form has a pair of secondary imines in the isoalloxazine ring.

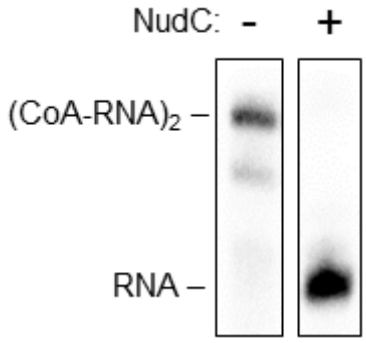


Figure 4.19 NudC Treatment of CoA-RNA.

Internally labeled CoA-RNA was treated with 4 μM enzyme for 30 min at 37 °C, indicating that both reduced and oxidized CoA-RNA were capable of being degraded by NudC.

NudC was tested for its ability to hydrolyze CoA-RNA as reported by Bird et al. (2016). Internally labeled CoA-RNA was incubated with 4 μM NudC for 30 minutes at 37 °C and run on 8% denaturing PAGE to visualize the shift in electrophoretic mobility. The removal of PhP from CoA-RNA should result in a shift similar to that of a 1 nt difference. While visualizing 1 nt differences is possible by PAGE, the hydrolysis of oxidized CoA-RNA is more easily observed based on the migration rate differences between the dimer RNA and the uncapped monomer (Fig. 4.19). The successful hydrolysis of oxidized CoA-RNA indicates that NudC is also capable of hydrolyzing oxidized CoA itself. This activity therefore corroborates the conclusions of Bird et al. (2016) in terms of the *in vitro* activity of NudC, and extends the previously characterized substrate range of NudC.

The role of NudC as a general coenzyme hydrolase seems evident in its similar activities in hydrolyzing the most common coenzymes CoA, FAD, and NAD. Given the overlapping specificities for each coenzyme, the discrimination

between them seems minor, and the binding of these substrates is likely due to the adenosine of each in addition to assisting non-specific hydrophobic interactions, therefore greatly reducing the activity against charged species such as oxidized NAD^+ (Frick and Bessman 1995). Given that intracellular CoA concentrations can range from 30 to 400 μM, NudC could operate in nutrient rich conditions as a CoA hydrolase near its optimum rate, which may have a large impact on cell survival given the versatility and importance of free CoA (Miller et al. 2007, Leonardi et al. 2005). By the same logic, NAD^+ would be cleaved at far below the optimum rate of NudC as well, given the high K_m in comparison to cellular concentrations of NAD, and even further below this rate when the ratio of free NAD is considered (Lin and Guarente 2003). As NudC has been demonstrated to hydrolyze CoA- and NAD-RNA, it seems logical that it would be involved in regulating these RNAs by degradation (Bird et al. 2016). The concentrations of CoA-RNA and NAD-RNA, however, are even farther below that of free CoA and NAD in the cell, and therefore would not be sufficient for NudC to operate at a rate required for immediate regulation assuming that the kinetic parameters for CoA and NAD are the same as those for CoA-RNA and NAD-RNA (Kowtoniuk et al. 2009, Chen et al. 2009). As such, the activity of steady-state concentrations NudC may act as a barrier to prevent the flux of such coenzyme-capped RNAs, maintaining their abundance below a certain point within the cell, and upon environmental stimuli is tuned by the increased expression of NudC itself. NudC has the third highest transcript abundance of the Nudix enzymes in *E. coli* with ~1300, however it falls below NudB (3200) and

NudH (3100) which are involved in the hydrolysis of oxidized dADP and di-adenosine polyphosphates, respectively (McLennan 2006). It's unclear whether the transcript abundance is correlated to active enzyme concentration within the cell, but the presence of such transcript abundance may support the hypothesis of the induced translation of NudC upon environmental stimuli to initiate gene regulation via coenzyme-RNA decapping.

Given the ability for NudC to act on a variety of adenosine-containing substrates, it seems unlikely that it is the sole enzyme responsible for the degradation of NADH, NAD^+, $FADH_2$, FAD, CoA, Acetyl-CoA, Succinyl-CoA, and their previously observed RNA conjugates. Such a one-size-fits-all mechanism of degradation would imply that distinct coenzyme-RNA conjugates exist as mechanistic redundancies and do not hold specific purposes. If all coenzyme-RNA conjugates hold the same function, then logically their degradation could be regulated by the same enzyme. This seems highly unlikely given the difference in mechanisms required for the installation of NAD and CoA as transcription initiators. While NAD can be readily incorporated as initiators into transcripts where A is the +1 nucleotide given an amenable promoter and yet-to-be-elucidated specific σ-factor, the same cannot be said for CoA (Bird et al. 2016). Transcription initiation is only possible with depCoA, the penultimate intermediate of the CoA synthesis pathway. While depCoA is present in the cell between the activities of CoaD (PPAT) and CoaE (depCoA kinase), its abundance is not readily detectable by any known method that distinguishes it from CoA and is therefore essentially unquantifiable. The amount of depCoA required to initiate

transcription would need to be high enough to compete with the abundant ATP in the cell as stated by recent work, however the intracellular concentrations of CoA itself have only been noted to increase up to 400 μM (Bird et al. 2016, Miller et al. 2007). Assuming this estimate includes both depCoA and CoA, and assuming depCoA only represents a small fraction of this population, it is unlikely that depCoA exists in concentrations that would allow transcription initiation. It therefore is unlikely that the single enzyme NudC would be responsible for regulating two separate coenzyme-RNA conjugates that are synthesized by presumably different pathways for presumably different reasons.

While certainly capable of hydrolyzing these RNA-conjugates *in vitro*, it remains to be seen through knockout mutation studies whether or not NudC is involved in the regulation of both *in vivo*. Regardless of its role as an RNA-conjugate regulator, NudC has been demonstrated capable of hydrolyzing all common diphosphate-containing coenzymes with estimated kinetic parameters that imply some level of activity in the cellular environment. Pending further investigations into its specific role *in vivo* as an RNA-conjugate decapper, NudC should be considered a general coenzyme hydrolase.

4.2.4 E. coli NudL

4.2.4.1 Cloning and Expression of SUMO-NudL

With all of the recent interest in Nudix hydrolases that came with the discovery of CoA- and NAD-RNA in 2009, the identification of a specific CoA hydrolase in *E. coli* has remained overlooked. Despite being identified by protein BLAST searches for both the CoA and Nudix box consensus motifs, the *E. coli*

gene NudL (YeaB) had never been characterized (McLennan 2006). Given the high likelihood that this enzyme would foster activity towards CoA and our successful cloning protocol, NudL was cloned, purified, and characterized as a specific CoA-hydrolyzing enzyme.

The NudL gene was amplified from the *E. coli* genome as previously described using primers designed to add flanking regions of 18 bp homologous to the pETite-nHis-SUMO-Kan vector (Fig. 4.20A). The PCR product was combined with linear pETite-nHis-SUMO-Kan in High-Control 10G cells for transformation and recombination, and Colony PCR analysis indicated 80% of screened transformants had successful generated the plasmid pETS-NudL (Fig. 4.20B). Significantly more positive transformants were noted when cloning NudL than NudC, despite their source from the same genome using the same high fidelity Q5 polymerase. After growing cultures for plasmid extraction, PCR confirmed a band near 729 bp that indicates the successful insertion of the NudL gene downstream of the SUMO fusion, and therefore successful generation of the pETite-SUMO-NudL plasmid (Fig. 4.20C).

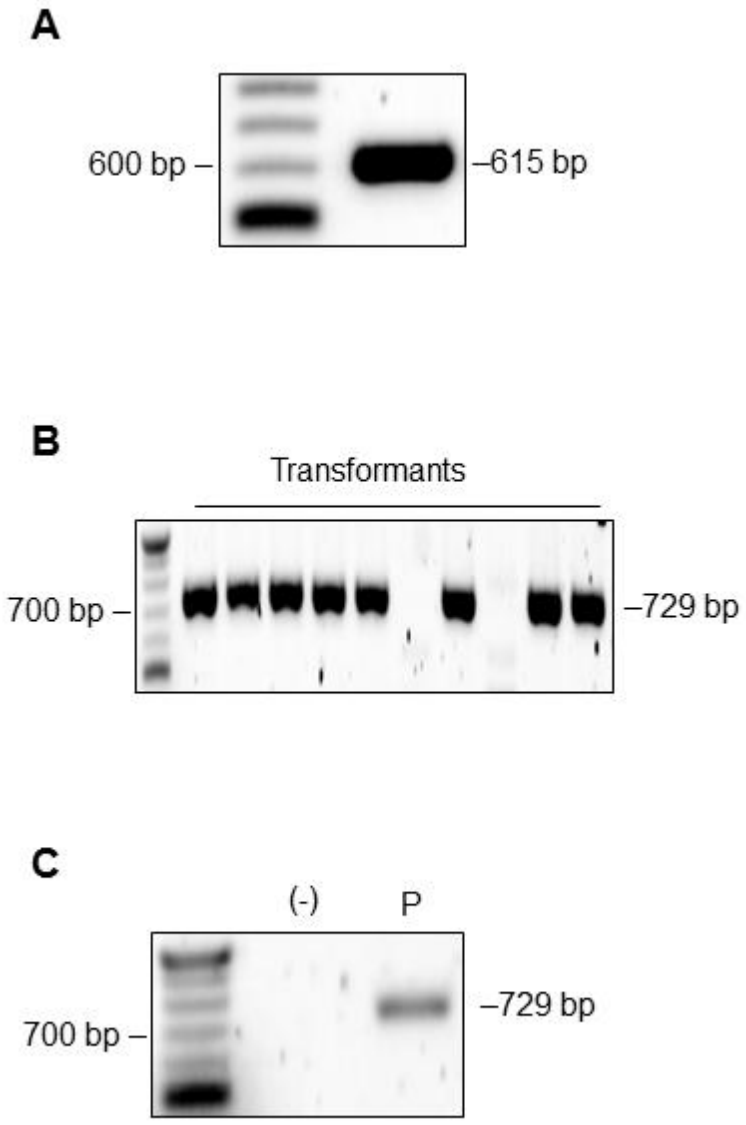


Figure 4.20 Cloning of SUMO-NudL.

A, Amplification of NudL gene from the *E. coli* genome by 30 cycles of PCR with primers to add 18 bp regions of homology to the pETite-nHis-SUMO-kan vector for recombination during transformation. B, Colony PCR screening of High-Control 10G transformants using general primers for the pETite vector. C, Verification of pETS-NudL plasmid extracted from High-Control 10G transformants by PCR with general primers for the pETite vector.

The PCR- verified plasmid was used for transformation of High-Control Bl21 (DE3) cells used for expression. Upon successful Colony PCR analysis of transformants, small scale expression reactions were carried out in 5 mL antibiotic media where cultures were grown at 37 °C until OD_{600} ~0.5, at which point expression was induced with the addition of 0.5 mM IPTG and allowed to express for 3 h at 37 °C. Each culture showed high expression of the 34 kD SUMO-NudL fusion protein (Fig. 4.21A), however solubility assays showed the vast majority of the protein was expressed insolubly (Fig. 4.21B). Expression conditions were repeated with induction temperature lowered to 16 °C, with decreasing inducer concentrations down to 50 μM IPTG, and reduced induction times, but soluble protein could not be obtained (Fig. 4.21C).

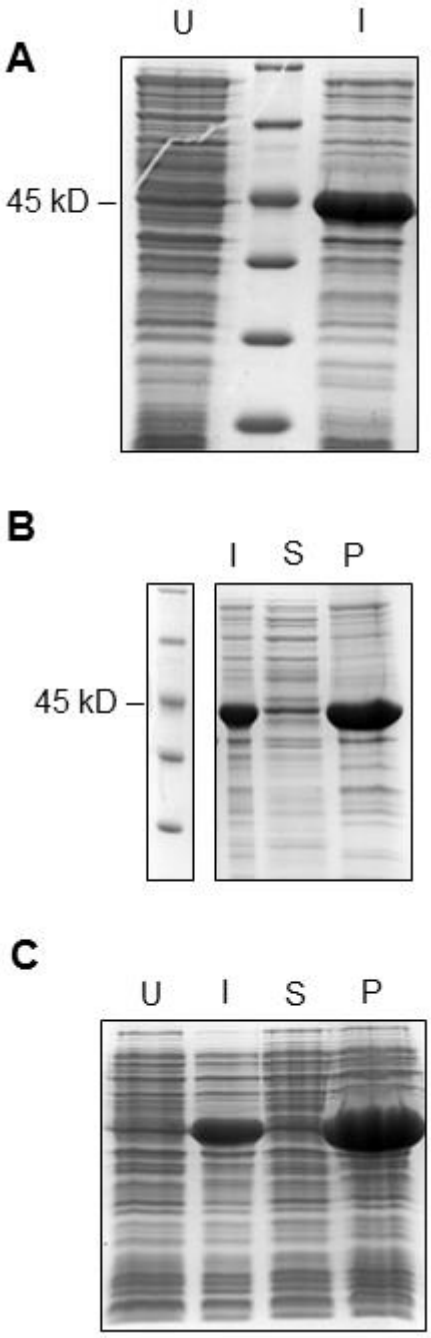


Figure 4.21 SUMO-NudL Protein Expression.

A, SDS-PAGE of small scale expression from High-Control Bl21 (DE3) transformants containing the pETS-NudL vector with induced (I) and uninduced (U) samples showing expression of the 34 kD SUMO-NudL fusion protein. B, SDS-PAGE of solubility assay from small scale expression with induced (I), soluble fraction (S), and insoluble fraction (P) samples

showing SUMO-NudL not expressed solubly. C, SDS-PAGE of larger scale expression induced at 16 °C with uninduced (U), induced (I), soluble (S), and pellet (P) samples indicating that protein is expressed insolubly.

4.2.4.2 Cloning and Expression of MBP-NudL

After careful consideration, it was decided that the T7 promoter was too strong and expression could not be tuned down to a level where the protein would stay soluble. Therefore NudL was prepared for expression in a system controlled by the Tac promoter using a maltose-binding protein fusion. Maltose-binding protein fusions have been used to increase the solubility of recombinant proteins due in part to its high solubility and to its large size of ~40 kD (Guan et al. 1988). In addition to increasing solubility of recombinant proteins, MBP-fusions are capable of purification by amylose resins which contain fewer steps to clean and prepare than typical IMAC resins.

The NudL gene was prepared for recombination into the pMBP-Parallel vector by PCR amplification from the *E. coli* genome using primers that would add flanking regions of 18 bp homologous to plasmid, remove the multiple-cloning site, and add a thrombin recognition site for cleavage of the fusion if necessary (Fig. 4.22A). The vector was prepared by PCR using primers designed to amplify the backbone to generate a linear stock as described above. The amplified products would share 18 bp of homology at the 3’ end of the ORF near the in-frame stop codon as well as just downstream of the MBP fusion, yielding a construct that contains the MBP fusion, thrombin cleavage site, and NudL under the control of the Tac promoter. PCR products were combined in DH5α cells for transformation and recombination, and Colony PCR screening showed that 50% of transformants had successfully generated the plasmid pEML

(Fig. 4.22B). Further PCR analysis of extracted plasmids with general primers confirmed that NudL had been inserted downstream of the MBP fusion (Fig. 4.22C).

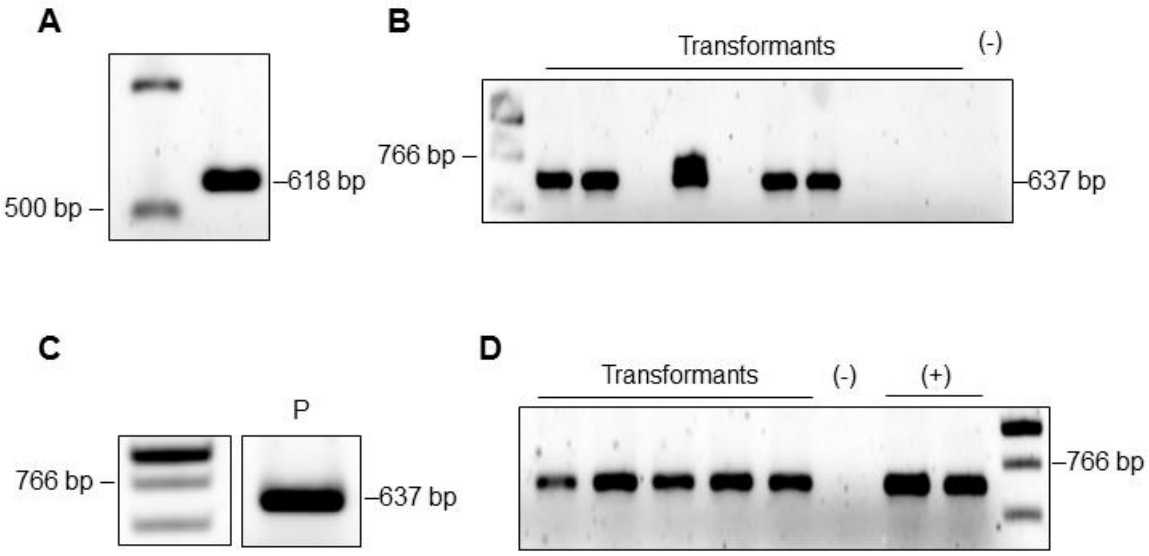


Figure 4.22 Cloning of MBP-NudL.

A, Amplification of NudL gene from *E. coli* genome by 30 cycles of PCR using primers to add 18 bp homology to pMBP-Parallel vector. B, Colony PCR of DH5α transformants using general primers for pMBP-Parallel that indicate the successful production of pEML.

PCR-verified plasmid pEML was used to transform High-Control Bl21 (DE3) cells (Fig. 4.23D). Cultures were grown for small scale expression in 5 mL of antibiotic media at 37 °C until OD_{600} ~0.5 when they were induced with the addition of 0.2 mM IPTG and allowed to express for 2 h at 37 °C. Analysis of small scale cultures by SDS-PAGE indicated good expression of the ~64 kD MBP-NudL fusion in all samples (Fig. 4.24A), therefore a culture was grown in larger scale with 0.5 L antibiotic media to OD_{600} of ~0.5 before induction with 0.2

mM IPTG and incubation at 16 °C overnight. The resulting protein was extracted in the soluble fraction and purified via amylose resin where it was eluted in the presence of 20 mM maltose (Fig. 4.23B). Eluted fractions were combined on an M50 spin column and washed repeatedly with 1X PDB to lower the salt concentration and remove the maltose before the addition of glycerol. Protein quantity was determined by absorbance at 280 nm using the calculated extinction coefficient of 93,280 M^{-1} cm^{-1} and molecular weight of 64,484 Da for a final yield of 4 mg L^{-1}. The relatively lower yield is expected from a non-T7 expression system, however the reduction in expression with the addition of the MBP fusion greatly helped the solubility of the protein.

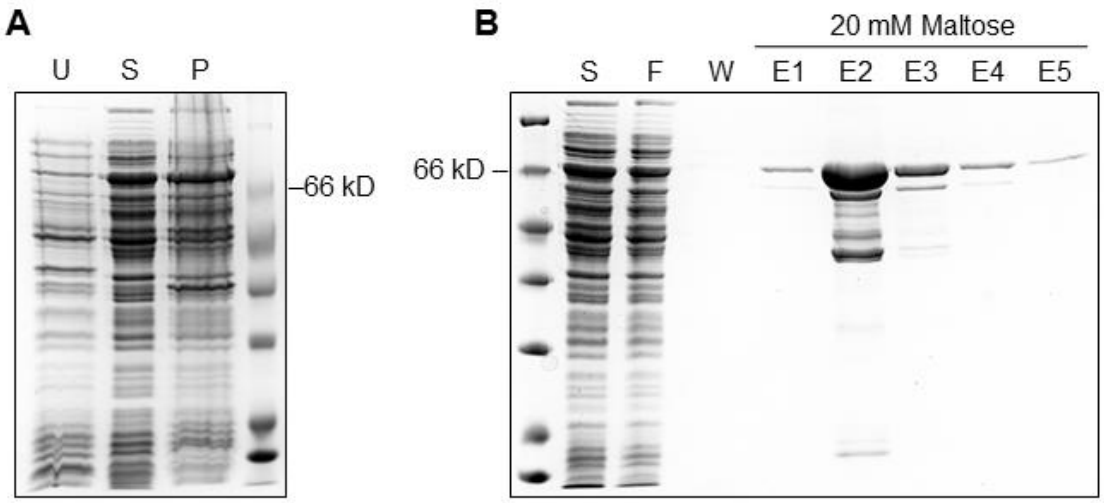


Figure 4.23 MBP-NudL Protein Expression.

A, Verification of plasmid (P) extracted from DH5α transformants by PCR with general primers for the pMBP-Parallel vector. B, Colony PCR of High-Control Bl21 (DE3) transformants carried out in batch with water (-) and two previously verified versions of pEML (+). E, SDS-PAGE of solubility assay containing uninduced (U), soluble (S) and insoluble lysate (P) samples to indicate 64 kD MBP-NudL protein is expressed in the soluble fraction. F, SDS-PAGE of MBP-NudL purification by amylose resin showing soluble lysate (S), column flow-through (F), column wash (W), and 1 mL fractions eluted with 20 mM maltose (E1-5).

4.2.4.3 Activity of NudL

The activity of NudL was tested in the presence of varying concentrations of CoA in 1X Nudix Buffer at 37 °C and visualized by HPLC (Fig 4.24C). NudL was determined to have a K_m of 0.51 mM for CoA with a K_{cat} of 0.008 s^{-1} and an enzyme efficiency of 0.015 mM^{-1} s^{-1} (Fig. 4.24D, Table 4.2). These values are all comparatively low to previously characterized Nudix CoAses from other organisms, in particular the K_{cat}/K_m of NudL is 2 orders of magnitude below both *M. musculus* Nudt7 and *C. elegans* NDX8 (AbdelRaheim and McLennan 2002, Gasmi and McLennan 2001). While both the mouse and worm enzymes are two

of the most efficient Nudix CoAses characterized, the turnover number of NudL is still significantly lower in comparison (Table 4.3). The half-saturation constant of NudL is double that of Nudt7 and NDX8, but it is still comparable to the cellular CoA concentrations of 400 μM expected in prolific, nutrient-rich conditions (Leonardi et al. 2005).

Table 4.3

Kinetic Constants of Nudix Hydrolases for CoA including NudL

Enzyme	K_m *(mM)*	K_{cat} *(s^{-1})*	K_{cat}/K_m *(mM^{-1} s^{-1})*
NudC	0.210	0.206	0.984
NudL	0.507	0.008	0.015
NDX8*	0.22	13.8	64
Nudt7**	0.240	3.8	16

*Data taken from AbdelRaheim et al. (2002). **Data taken from Gasmi et al. (2001)

Indeed, the kinetic constants for NudC were also calculated lower than the specific CoAses from worm and mouse, but the K_m values of all three are comparable. If cellular concentrations reach to 400 μM in nutrient rich conditions, NudC would be operating at its maximum rate and opposing the metabolic capabilities allowed by the environment. In other words, when *E. coli* is in high glucose and can achieve maximum growth rate through the use of CoA, NudC would be active in removing this indispensable metabolic component, and therefore NudC must be kept under some translational control or an active

repression. NudL, on the other hand, contains the CoA consensus sequence shared by NDX8, Nudt7 and others, and is specific for CoA with a lack of activity towards NAD or FAD. The transcript abundance of NudL is only ~15% that of NudB, and ~36% of NudC, indicating that significantly fewer cell resources are devoted to maintaining the amount of mRNA present (McLennan 2006). As previously stated, it is unclear whether the transcript abundance correlates to the number of actual proteins present in the cell, therefore transcript abundance should be taken lightly when concluding protein activity within the cell.

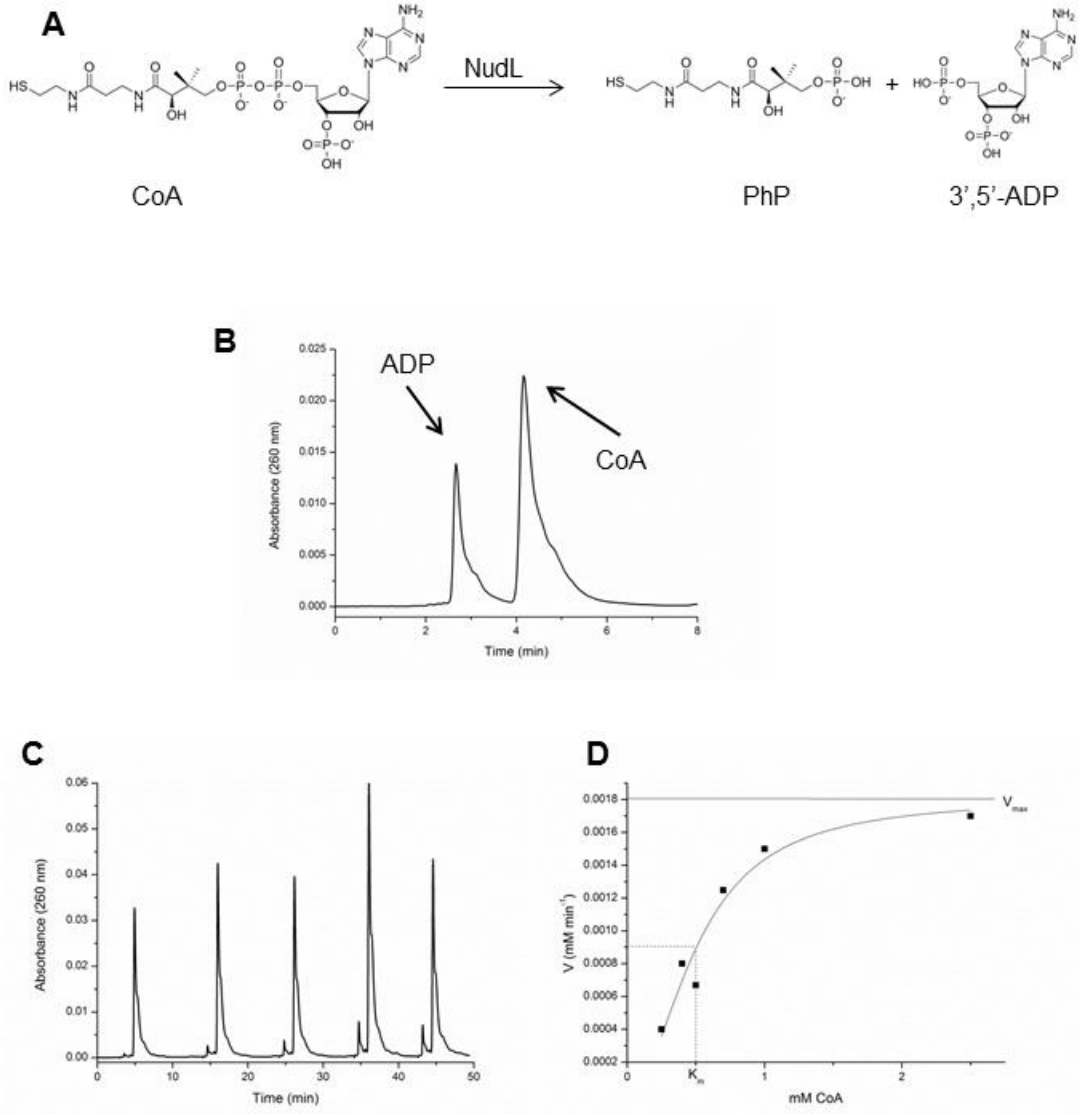


Figure 4.24 Activity and Kinetics of NudL with CoA.

A, Scheme of NudL hydrolysis of CoA into 4'-phosphopantetheine (PhP) and 3', 5'-adenosine diphosphate (3', 5'-ADP). B, HPLC chromatogram showing the separation of products (ADP) from reactant CoA due to NudL hydrolysis. C, HPLC chromatogram showing progress of CoA hydrolysis reaction over time. D, Michaelis-Menten plot of kinetic data obtained from HPLC monitoring of CoA hydrolysis by NudL, where solid horizontal line represents V_{max} and dashed line represents the half-saturation constant K_m.

The low activity assumed from the kinetic constants of NudL may be advantageous for the cell, as a highly active CoA hydrolase could be problematic

due to the removal of a key metabolic coenzyme. Even in the event of high CoA concentrations, the enzyme's efficiency is still an order of magnitude below that of NudC, implying that the removal of the highly important CoA is not at as high of a risk. The stark difference in K_m and K_{cat} values between NudL and its eukaryotic counterparts may be explained by the evolution of the eukaryote itself. Simply put, the *E. coli* enzyme must function in the cytosol in the presence of the CoA pool that is so vital to metabolism, whereas the *C. elegans*, *S. cerevisiae*, and *M. musculus* versions of the Nudix CoAse are separated from their respective cellular CoA pools by the presence of organelles. Each of the eukaryotic enzymes have been shown to localize to the peroxisome as a result of a signal peptide, whereas the major pool of CoA in eukaryotes is found in the mitochondria (Cartwright et al. 2000, Gasmi and McLennan 2001, AbdelRaheim and McLennan 2002). Therefore the eukaryotic enzymes have no selective disadvantage to high enzyme efficiencies, but such an enzyme in bacteria could be detrimental to survival. While NudC has an estimated enzyme efficiency two orders of magnitude above NudL, it is still one order of magnitude below the eukaryotic Nudix CoAses (Table 4.3).

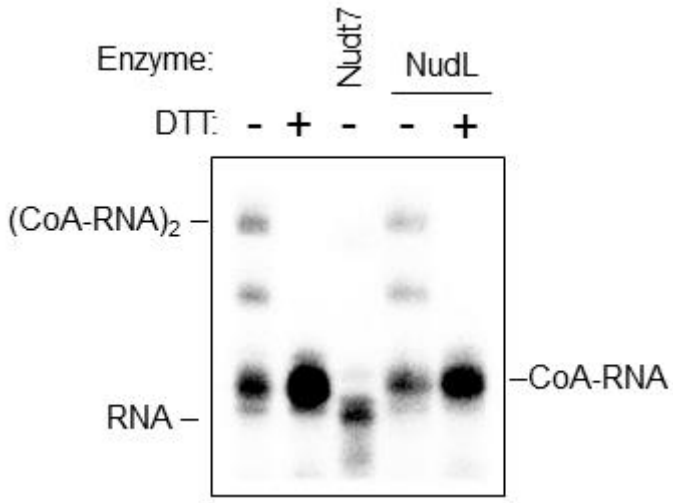


Figure 4.25 NudL Treatment of CoA-RNA

Internally labeled CoA-RNA was treated with 4 μM enzyme and 100 mM DTT. NudL activity could not be observed with oxidized or reduced CoA-RNA in the 1 h incubation time at 37 °C.

The activity of NudL against CoA-RNA was tested the same as experiments with NudC. Internally labeled CoA-RNA was incubated with 4 μM enzyme for 1 h at 37 °C and resolved on 8% denaturing PAGE to visualize the removal of PhP and the resulting electrophoretic mobility shift (Fig. 4.25). The altered migration rate is approximately equal to that resulting from a 1 nt difference, and can easily be visualized using this method as shown by the treatment of CoA-RNA with Nudt7. The addition of NudL to the CoA-RNA did not appear to alter the migration rate in either the absence or presence of a reductant, as both lanes appear similar to controls without added enzyme. The inability to visualize CoA-RNA decapping by NudL may be caused by the lack of NudL to process CoA-RNA, or it could be a result of NudL's drastically lower catalytic efficiency compared to NudC and Nudt7. Both NudC and Nudt7 were

capable of hydrolyzing visible amounts of CoA-RNA, however these enzymes have 20-fold and 400-fold higher catalytic efficiencies, respectively, in addition to K_m values half that of NudL. Therefore the hydrolysis of CoA-RNA by NudL may be occurring at a rate so slow that products are not detectable. To account for the 20-fold difference in catalytic efficiency between NudL and NudC, CoA-RNA would need to be incubated approximately 10 hours to achieve the same level of decapping as observed by NudC. Experiments were therefore repeated extending the time to 4 hours incubation, but still no detectable CoA-RNA was hydrolyzed by NudL.

The possibility of the MBP fusion interfering with the activity of NudL against CoA-RNA cannot be ignored. While still capable of cleaving CoA without any observable issue with the fusion protein attached, the large nature of MBP may sterically hinder the insertion of CoA-RNA into the active site of NudL, causing the MBP-fusion enzyme to appear inactive against this substrate. The MBP-NudL protein was therefore treated with thrombin overnight at both room temperature and 37 °C to remove the MBP fusion. The resulting crude protein was used in similar experiments with CoA-RNA without any change in activity. Strangely, NudL was also found incapable of cleaving CoA at an equivalent rate as its pre-processed counterpart MBP-NudL, indicating that the protease digestion of the enzyme had resulted in either denaturation of the protein, or off-target proteolysis to reduce the efficiency of the enzyme.

NudL was not observed to decap CoA-RNA, despite its ability to hydrolyze CoA derivatives such as depCoA, oxidized CoA, and oxidized depCoA at

comparatively slower rates than CoA itself. Given the size of molecules such as oxidized CoA, it seems that NudL should be capable of hydrolyzing 15mer CoA-RNA as well. The issue is likely with the concentration of CoA-RNA used in the analysis. Experiments with CoA derivatives were carried out with 1 mM substrate which is far above the calculated K_m for CoA when the enzyme would be operating near its optimum rate, but CoA-RNA could never be used in such high concentrations due to lack of materials. The estimated amount of 15mer CoA-RNA used in experiments is closer to 1 μM, therefore causing the rate of NudL to be significantly slower. Considering these parameters in addition to the significantly slower turnover number calculated for NudL, it seems unlikely that CoA-RNA hydrolysis by NudL can be visualized without a significant increase in enzyme concentration.

4.3 Conclusions

The cloning of four Nudix enzymes capable of hydrolyzing CoA in 8 different iterations was successfully carried out in this study. Solubility issues limited the success of purification of each of these enzymes. These issues were overcome with the latter three enzymes Nudt7, NudC, and NudL, allowing the functional characterization of each for the assessment of their potential use in a CoA-RNA capture protocol. NudL was determined to be less plausible of an enzyme to use for such a protocol as NudC or Nudt7 which both bear superior kinetic characteristics. Nudt7 and its ability to cleave CoA-thioesters and biotinylated-CoA-RNA, in addition to its high rate of hydrolysis, appears to be the best candidate for the task of CoA-RNA hydrolysis from resin. NudC also bears

the capability for this task, however the specificity of its activity comes into question. The purpose of CoA-RNA capture is selective release of CoA-RNA by specific enzyme activity. Such specificity is displayed by Nudt7, but not NudC (Gasmi and McLennan 2001, Reilly et al. 2008, Frick and Bessman 1995). Given the conditions of the CoA-RNA capture protocol to be discussed in a later section, the potential for off-target cleavage by a general hydrolysis enzyme such as NudC is problematic. The NAD-RNA capture protocol was very specific to capture only oxidized NAD, and yet the supplemental data the investigators provided showed very dirty results that caused their analysis to be limited to RNAs isolated of specific sizes (Cahova et al. 2015). The most stringent conditions are required in order to increase the purity and depth of the CoA-RNA capturing, and therefore the inclusion of NudC should be carefully considered before proceeding.

This study also describes the novel characterization of two Nudix enzymes capable of hydrolyzing CoA, one of which had never been studied at all. It is unclear why *E. coli* would produce the two enzymes NudL and NudC with overlapping functions unless each was intended for use in specific scenarios. NudL may be strictly involved in the slow and steady metabolism of CoA itself whereas NudC may be used for a more general purpose of coenzyme degradation to be specifically induced for gene regulation. At the current calculated kinetic constants, neither enzyme is capable of quickly turning over CoA-conjugated RNA given its estimated cellular concentration (Kowtoniuk et al. 2009). The rate of CoA-RNA hydrolysis would certainly increase with protein

concentration, leaving the possibility that NudL expression could be tuned by environmental stimuli to induce the degradation of CoA.

Given the specificity of a motif present in NudL for CoA, CoA-RNA should be considered its substrate as well. Experiments carried out with NudL and internally labeled CoA-RNA, however, were unable to demonstrate this hydrolysis. It is possible that the position and size of the MBP fusion on NudL interferes with both the kinetics of the enzyme and its capability to hydrolyze CoA-RNA, but several attempts to remove it by treatment with thrombin as designed resulted in an inactive protein that was incapable of hydrolyzing CoA itself at detectable rates. If the MBP fusion did not abrogate the ability of NudL to hydrolyze CoA, the removal of the fusion should show a similar amount of activity, if not better. Therefore subjecting the protein to thrombin may have resulted in off-target degradation of the protein. The use of thrombin for fusion removal is a popular technique, but off-target proteolysis is a known characteristic of this protease (Jenny et al. 2003).

Considering the slow rate of CoA hydrolysis by NudL, it is possible that CoA-RNA can still be turned over at a rate incapable of detection by the current methods employed. The removal of PhP by NudL hydrolysis of CoA-RNA would result in an electrophoretic mobility shift equivalent to ~1 nt. While this can easily be visualized by PAGE, a significant fraction of the CoA-RNA needs to be hydrolyzed in order for the shift to appear as a band and not simply a smear. These assays can easily be used to view CoA-RNA hydrolysis by both NudC and Nudt7 due to their increased activity in comparison to NudL. The assumption that

hydrolysis of CoA-RNA by NudL occurs at a comparably slower rate than observed for CoA reinforces the theory that the slow NudL gives a selective advantage to *E. coli*. The high K_m for CoA calculated for NudL is near the observed upper limits of CoA flux by other investigators, implying that the enzyme would not be active in the cell at a rate that would hinder metabolism even in the most prolific of conditions (Leonardi et al. 2005). CoA-RNA is estimated to be significantly lower in concentration than CoA at around 8 fmol μg^{-1}, and therefore if the activities of CoA and CoA-RNA were assumed to be equivalent, NudL would be comparably less active against CoA-RNA in the cell (Kowtoniuk et al. 2009). Therefore the activity of NudL within *E. coli* would likely not interfere with the function of CoA-RNA through transient degradation as hypothesized with NudC. An enzyme such as NudL would be reasonable as a tunable control for CoA-RNA levels, and such a situation would allow the cell to specifically control CoA-RNA turnover without affecting NAD-RNA levels that would result from NudC activity.

The nonspecific hydrolysis and substrate ambiguity displayed by NudC insinuates that other enzymes could be capable of cleaving CoA-RNA without the specificity associated with its primary sequence. Such a topic has been discussed in recent reviews, as Nudix enzymes have been characterized with a multitude of substrates (McLennan 2013). The fact remains that activities discovered by *in vitro* methods do not necessarily reflect the native activity of the enzyme *in vivo*. Surely, NudC is capable of catalyzing a vast array of nucleotide-derived molecules, but does it actually exist in such abundance within the cell

that it is capable of such activity? Reasonably the answer is no. If NudC was actively carrying out all of its *in vitro* characterized hydrolyses within the cell, then no species expressing it would be able to survive. Of the 13 identified Nudix hydrolases in *E. coli*, 4 are still uncharacterized including NudL (McLennan 2006). These remaining enzymes YffH, YfaO, and YfcD likely display some NudC-like nonspecific activity considering that no substrate specific motifs have been identified within them aside from the NDP-sugar affinity predicted with YffH. Other possible proteins capable of CoA-RNA hydrolysis following the mechanisms of NudL and NudC would maintain pyrophosphohydrolase activity, but not necessarily CoA specificity. Protein sequence analysis of related sequences to NudL and NudC spotlights a variety of NTP pyrophosphohydrolases that are uncharacterized in *E. coli* such as predicted CTP pyrophosphohydrolase, RNA pyrophosphohydrolase, and Pyrimidine Deoxynucleoside Triphosphate pyrophosphohydrolase. Some of these results, especially those annotated for RNA pyrophosphohydrolase activity, seem to be promising candidates for CoA-RNA hydrolysis activity. Unfortunately, none have been studied to date. Furthermore, the Nudix superfamily itself is a subset of the much larger αβ protein family that is characterized by the structural motif of antiparallel β-sheets with segregated α and β regions. This αβ family consists of over 376 different superfamilies according to the Superfamily v1.75 annotation algorithm (http://www.supfam.org/), including proteinases, nucleases, RNase A-like, and DNA-binding domain containing proteins based on that characteristic structure.

With such an abundance of family members capable of binding nucleic acid and hydrolyzing phosphodiester or phosphoanhydride bonds, the possibility exists that many other enzymes could be capable of nonspecifically metabolizing CoA-RNA. Members of the nuclease superfamily, for example, would be quite capable of degrading CoA-RNA if they showed endonuclease or 3'-5' exonuclease activity. Enzymes with 3'-5' exonuclease activity would be able to degrade the CoA-RNA and release either depCoA or CoA itself to be recycled into metabolism or the CoA pathway. There is no reason to discount this as a pathway for CoA-RNA degradation, and such a 3'-5' exo pathway is certainly a possibility for any RNA found within the cell. The problem is with the nonspecificity of the 3'-5' exo pathway. The CoA cap has no influence on this pathway, but it does alter the 5'-3' pathway initiated by RppH to decap triphosphate RNA. It stands to reason that if there is a specific pathway to create such a molecule, there too should be a pathway to destroy it.

The only experimentally or computationally characterized enzymes capable of the destruction of CoA by the removal of PhP are Nudix enzymes discussed above, the enzyme phosphopantetheine adenylyl transferase (PPAT, CoaD) from the CoA synthetic pathway, and the acyl-carrier proteins involved in acyl transfer. The ACPs are characterized in lipid and polyketide synthesis where they acquire PhP to carry out acyl-transfer reactions. While this mechanism involves the use of CoA, it is unknown whether these proteins can interact with CoA-RNA. The enzyme CoaD is another possible candidate to carry out CoA-RNA hydrolysis, however it has been characterized to catalyze the reverse

reaction of phosphopantetheine adenylation that requires depCoA as a substrate. CoA-RNA, incidentally, does not contain a depCoA group due to the presence of the 3'phosphate connecting it to the remaining RNA residues, and CoaD has been shown to be incapable of this activity with CoA (Geerlof et al. 1999). Therefore it is unlikely that CoaD is also capable of CoA-RNA hydrolysis. Just as NudC is capable of CoA-RNA hydrolysis without specificity, it's possible that other proteins unearthed in the structural and sequence homology analysis could carry out this mechanism albeit without specificity for CoA-RNA. These enzymes will require further study and characterization to determine their roles.

Speculation in the past decade of the roles of different Nudix enzymes and coenzyme-conjugated RNAs has led to an interesting set of conclusions. While Nudix hydrolases are well-characterized in many organisms, it seems the field of study shifted away from physical experiments and brought into focus the wealth of genomic data that came with the advent of Next Generation Sequencing technology. Even upon the discovery of CoA-RNA, the mechanism for its generation in addition to its degradation have been argued in both logical and illogical manners, yet no specific data has been presented as evidence that could be used to elucidate a reasonable model.

The argument that CoA can be installed as an initiator of transcription given the correct concentrations is simply wrong (Bird et al. 2016). Not only would CoA not exist in the specified concentrations to outcompete ATP for initiation *in vivo*, but CoA itself is incapable of extension via RNA polymerase due to the presence of a 3' phosphate group. Kowtoniuk et al. (2009) implied this

point upon their initial discovery of CoA-RNA when they could not reproduce CoA-RNA using *E. coli* RNA polymerase for *in vitro* transcription in the presence of depCoA and promoters that encode A as the +1 nucleotide. Indeed, Bird et al. (2016) could show this reaction to succeed *in vitro* using slightly different concentrations, different promoters, and added σ-factors, however they concluded this was the likely pathway of incorporation *in vivo*. In addition peculiar conclusions about general coenzyme incorporation into RNA, Bird et al. (2016) looked to NudC as the explanation for their idea of a sole decapping enzyme in *E. coli*. While this enzyme was shown to cleave NAD and CoA *in vitro*, the *in vivo* experiments carried out cannot be used to infer any conclusion in regards to the process within the cell. The investigators isolated and identified RNA by Northern blot from *E. coli* that had been transformed with a plasmid encoding a set sequence from a promoter they had reasoned would yield coenzyme-capped RNA. The isolated RNA was treated with NudC to show a gel shift, and the conclusion was made that NudC responsible for decapping within the cell. Not only do these experiments not provide evidence for one mechanism of coenzyme installation over another, but they do not indicate what was actually hydrolyzed from the 5’ end of the RNA. Looking at the original characterization of NudC from the mid-1990s, the enzyme has long been known to be capable of hydrolyzing a wide variety of substrates based on the presence of adenosine somewhere in the molecule (Frick and Bessman 1995). Amidst all of these years and different studies, CoA hydrolysis in *E. coli* remained uncharacterized.

This study represents the first characterization of an *E. coli* enzyme capable of CoA hydrolysis, and identifies two possible candidates to fulfill the long uncharacterized metabolic role of CoA degradation. Given their different efficiencies and half-saturation constants for CoA, it is likely the two enzymes NudC and NudL fulfill different roles in the cell, for example to control the CoA turnover in log and stationary growth phases separately. The substrate ambiguity of NudC, however, is troubling. It's unclear whether the variety of identified NudC substrates would actually be hydrolyzed in the cell. If NudC is present in the cell with 1 mM NADH, 400 μM CoA, and 10 μM FAD, it will actively hydrolyze all three and could substantially hinder the cellular metabolism (Lin and Guarente 2003, Leonardi et al. 2005, Louie et al. 2003). In such a situation, NudC would serve well as an inducible facilitator of the transition to stationary growth, where metabolism slows to a steady-state. Both enzymes require more study in order to fully understand their purposes. One thing that seems clear based on the kinetic data presented above is that, in order for either of these enzymes to serve as the active cellular hydrolase against such a small population of CoA-RNA, the enzyme concentration within the cell would need to be increased, indicating a regulatory mechanism responsible for controlling the translation of the respective mRNAs.

In order to ascertain the true role of these enzymes within the cell, *in vivo* assays will need to be carried out in the future. Knockout mutants of NudC and NudL, in addition to other possible candidates, will need to be characterized for differences in abundance and quality of CoA-RNA captured. In addition, these

knockout mutants can be supplemented with plasmids expressing the knocked out genes under the control of the arabinose operon for a tunable control of protein level similar to those done in previous investigations (Balibar et al. 2011). With these mutants, the changes in CoA-RNA abundance can be measured as a function of the tunable levels of NudC or NudL during growth. Such experiments would prove more directly the roles of these enzymes in CoA-RNA metabolism as well as CoA turnover within cellular environments.

CHAPTER V - DEVELOPMENT OF A SPECIFIC COA-RNA CAPTURE PROTOCOL

5.1 Introduction

The discovery of bacterial coenzyme-RNA conjugates has sparked new ideas regarding the bacterial transcription. The previously accepted prokaryotic transcription system simply polymerized nucleotide triphosphates (NTPs) to leave a 5'-triphosphate group at the terminus of all mRNA. This was as an easily understandable system. It did not bear the complexity of the eukaryotic system that includes 3' modifications via polyadenylation, or 5' modifications such as installment of a 7-methyl-guanosine (7mG) cap, or any of the protein interactions that go along with such modification. As such, the bacterial system was a reflection of the bacterial domain itself: simple and stripped down in comparison to its eukaryotic counterpart. Bacterial systems in general served as the fundamental building blocks for metabolic and systematic cellular processes that most complex systems in higher organisms were modeled after. As new aspects of RNA biology in higher organisms were discovered, the discovery of simpler versions in bacteria followed. One example is the discovery of eukaryotic microRNA (miRNA) transcribed from intron sequences in eukaryotic genomes that is responsible for specific post-transcriptional gene regulation that was followed by the discovery of bacterial small noncoding RNAs (sRNAs) that carry out the same function with fewer steps (Ha and Kim 2014, Bandyra et al. 2012). The same could previously be said about bacterial transcription. Bacterial mRNA are transcribed with a 5'-triphosphate at its terminus that serves to protect the

RNA from transient degradation by RNase E, and systems have been described that induce degradation via removal of the triphosphate group (Bandyra and Luisi 2013). Even with the addition of sRNA regulation, bacterial transcription and translation was a process of bare essentials without any excess RNA modification required for function. The eukaryotic system, while still maintaining the core values displayed in prokaryotes, has additional levels of complexity regarding both modification and RNA-protein interactions in order for the RNA to serve its intended function.

The viewpoint of bacteria as a simple model system was cracked with the discovery of NAD- and CoA-RNA (Kowtoniuk et al. 2009, Chen et al. 2009). For all that can be said about the complexity of eukaryotic RNA, there is generally one accepted cap structure of 7mG, and it is a modified form of a nucleotide that is not known to function in any other way within the cell. The bacterial system as of 2009 includes two separate 5'-modifiers, both of which not only are involved in a myriad of cellular processes, but themselves are indispensable to survival. CoA, in particular, participates in over 500 known reactions including those involved in fatty acid metabolism, glucose metabolism, and as a source for phosphopantetheine for carrier proteins (Leonardi et al. 2005, Kanehisa et al. 2008). NAD, on the other hand, is known for its roles in redox reactions in metabolism, adenylation, ADP-ribosylation, and DNA repair (Lin and Guarente 2003, Lin 2007, Garten et al. 2009). The discovery of these two coenzyme-RNA conjugates brought many questions to light. What is the function of these two coenzyme-RNA conjugates, and both coenzymes function the same way with

RNA? How are these coenzyme-RNA conjugates made, and are they made the same way? Do these coenzyme-RNA conjugates act in a similar way to the eukaryotic 7mG cap? How does the cell regulate the installation and removal of these coenzymes? The answers to these questions all center around a common theme that, for once, bacteria may no longer be considered a simple model system from which all other processes in life are expanded.

A group of investigators was able to shed light on one of these questions with the development of a specific NAD-RNA capture protocol (Cahova et al. 2015). The captured NAD-RNA were identified through Next Generation Sequencing and determined to be mostly sRNA of approximately 300 nucleotides. Cahova et al. (2015) extended their observations that the NAD group serves as an extra level of protection for these RNAs against RppH processing that leads to RNase E degradation, and that they could instead be processed by the Nudix enzyme NudC that was characterized as an NADH-hydrolase (Frick and Bessman 1995). If NAD was being installed as a form of degradation protection for sRNA, then how was this occurring? It is possible for NAD, FAD, and depCoA all to be installed to RNA as an initiator using the T7 RNA Polymerase (Huang 2003a). However it is also possible for these coenzymes to be generated via ribozyme activity from their respective substrates NMN, FMN, and PhP with an ATP molecule (Huang et al. 2000). As Kowtoniuk et al. (2009) noted, the incorporation of depCoA into RNA using the *E. coli* RNA polymerase is problematic, as their attempts at recreating CoA-RNA *in vitro* with such a system failed. Another group of investigators attempted to point out that

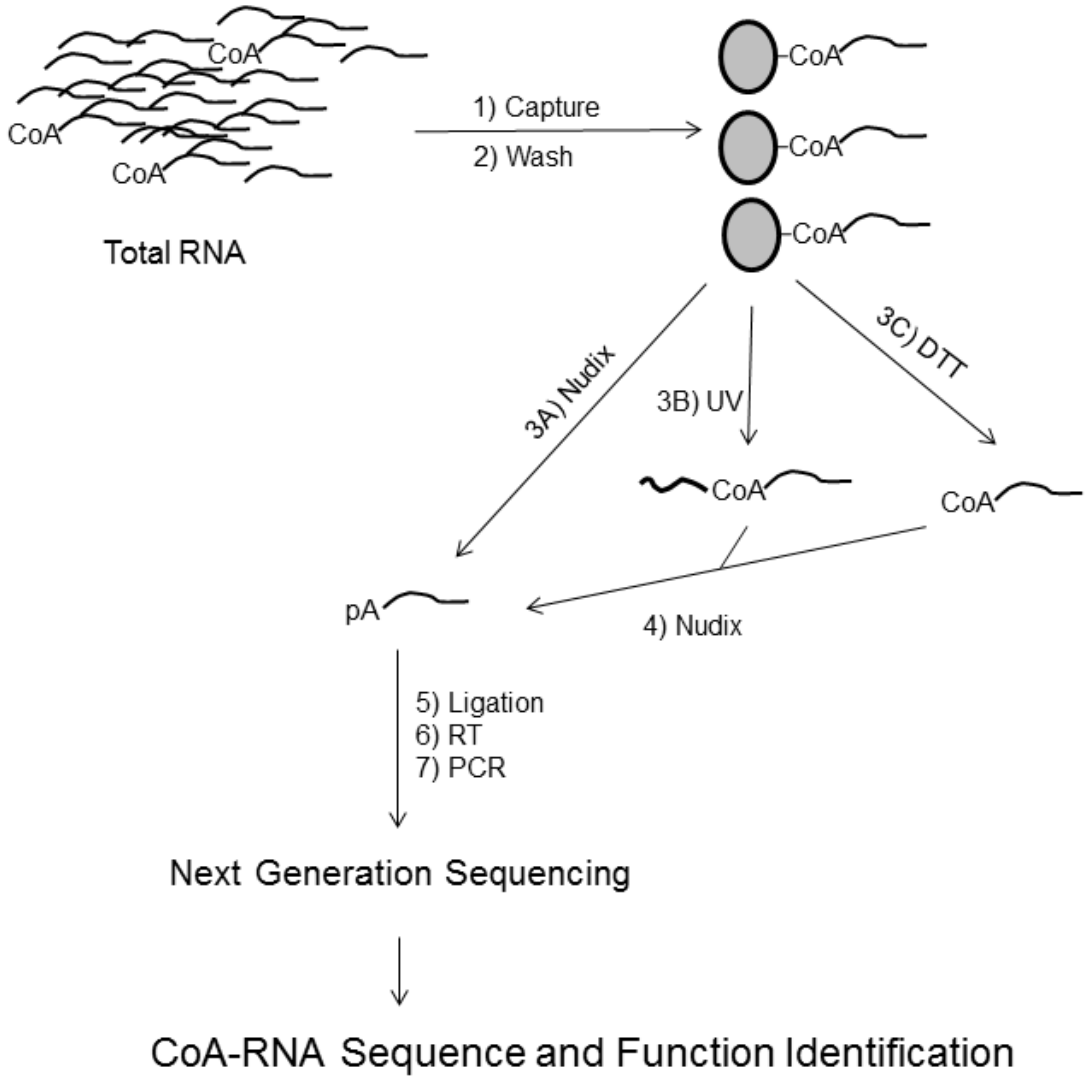
given the right ratios of NAD and depCoA to ATP, initiation with either could be achieved by outcompeting ATP for the +1 site of transcription (Bird et al. 2016). These results *in vitro* may be promising, however the *in vivo* concentrations of CoA, depCoA, and NAD are far from equivalent to each other, and in many cases far below that of ATP such that the ratio of coenzyme to ATP to achieve optimal transcription initiation of a coenzyme would not occur (Lin and Guarente 2003, Miller et al. 2007). While the concentrations of NAD could reach near 1 mM to facilitate such incorporation as described by Bird et al. (2016), CoA has only been known to increase to 400 μM in glucose-rich conditions (Leonardo et al. 1996, Miller et al. 2007). It should be understood that CoA is not capable of transcription initiation due to the presence of a 3'-phosphate group, and only its synthetic precursor depCoA can undergo such a reaction. DepCoA, being a synthetic intermediate to CoA, has never previously been quantified within the cells, and is indistinguishable from CoA in quantification methods. Regardless of the possibility that NAD concentrations reach high enough for transcription initiation to occur, the mechanism proposed then is that of competition, indicating a non-specific or process of capping RNA. Such a transient mechanism is not possible for CoA-RNA due to cellular concentrations of depCoA, so then why would one coenzyme-RNA conjugate be coincidental, but another be more pointed and require some specific machinery? The actual mechanisms of installment of these coenzymes are unknown. Everything at this point is speculation. The investigators previously published in this area all seem to believe transcription initiation to be the answer, regardless of the points raised

above. While certain promoters have been identified that initiate with an adenosine at the +1 region and allow the incorporation of NAD, NADH, and CoA as initiators *in vitro*, there is still no evidence of transcription initiation by coenzymes *in vivo* (Bird et al. 2016).

The idea that NAD-RNA and CoA-RNA should be considered different classes of RNA modifiers is more than simply the inability for one of them to transiently be incorporated into RNA as an initiator. The CoA-RNA conjugates discovered are unique. While only one type of NAD-RNA was discovered in the form of oxidized NAD^+, four different iterations of CoA-RNA were isolated (Chen et al. 2009, Kowtoniuk et al. 2009). In addition to CoA itself, the three separate thioesters of acetyl-CoA, succinyl-CoA, and methylmalonyl-CoA were discovered conjugated to RNA. Methylmalonyl-CoA is particularly interesting because it is incapable of being synthesized by *E. coli*, and therefore was only isolated from *S. venezuelae* cultures. The appearance of each these thioesters conjugated to RNA not only illustrates the diversity of CoA-related compounds available in different species, but it hints at the distinction of their functions from other coenzyme-RNA conjugates. It's interesting to think that CoA-RNA behaves as a contemporary version of a pre-protein enzyme, where the RNA is a stabilizing factor or a structural scaffold that facilitates the activity of CoA in acylation reactions (White 1976). Could the CoA conjugated to RNA therefore serve to not only as protection from RNA degradation as the NAD adduct was proposed by Cahova et al. (2016), but also still function in its acylation duties? Furthermore, are there sequence elements within the RNA that specifically direct CoA to form

these thioesters? Furthermore, do any of these sequences have similarities to those of NAD-RNA?

None of the questions presented by the discovery of CoA- and NAD-RNA can be completely answered without more data. The discovery of CoA-RNA will likely bring up more questions, but it is absolutely necessary to begin uncovering the relationships between these newly discovered coenzyme-RNA conjugates. To further explore the functions of the CoA thioester-RNA conjugates, a specific capturing protocol is required to distinguish between these and CoA-RNA. Here we describe the development of a specific CoA-RNA capture protocol to distinguish CoA-RNA from thioester-CoA-RNA, so that the RNA can be prepared for sequencing for its identification and the subsequent elucidation of its function (Scheme 5.1).



Scheme 5.1 Overall Capture Protocol.

The capture of CoA-RNA via resin conjugation (1) is essential to purification, as well as the extensive washing (2) to remove nonspecifically bound RNAs that would later show up as false positives in the sequence analysis. Removal of the CoA from the resin can occur by enzyme (3A), UV light (3B), or reduction (3C) depending on the resin used, but ultimately Nudix digestion (4) is required to generate the 5'-pRNA capable of ligation (5) and preparation of the NGS library (6, 7).

5.2 Results and Discussion

5.2.1 Capture of CoA-RNA by TGSIA

The initial plan for capturing CoA-RNA was to utilize polyacrylamide resin derivatized with a terminal iodo group. The free thiol of CoA would react with the iodo group in a nucleophilic substitution reaction to form a stable thioether bond (Fig. 5.1A). The commercially available resins containing iodo groups or amines that can be functionalized with succinimidyl iodoacetate (SIA) were obtained for this purpose. The iodo group is stable at acidic temperatures, but as the pH increases it releases from the resin forming I^-. Because the CoA would need an ionized thiol for the nucleophilic substitution to occur, the pH would need to be acid enough to prevent the non-specific loss of iodo from the resin, while still being basic enough to facilitate the nucleophilic substitution to form the thioether bond. The pH also would need to remain at or below 8 in order to maintain the integrity of RNA samples prone to alkaline hydrolysis due to the presence of their 2'-OH group.

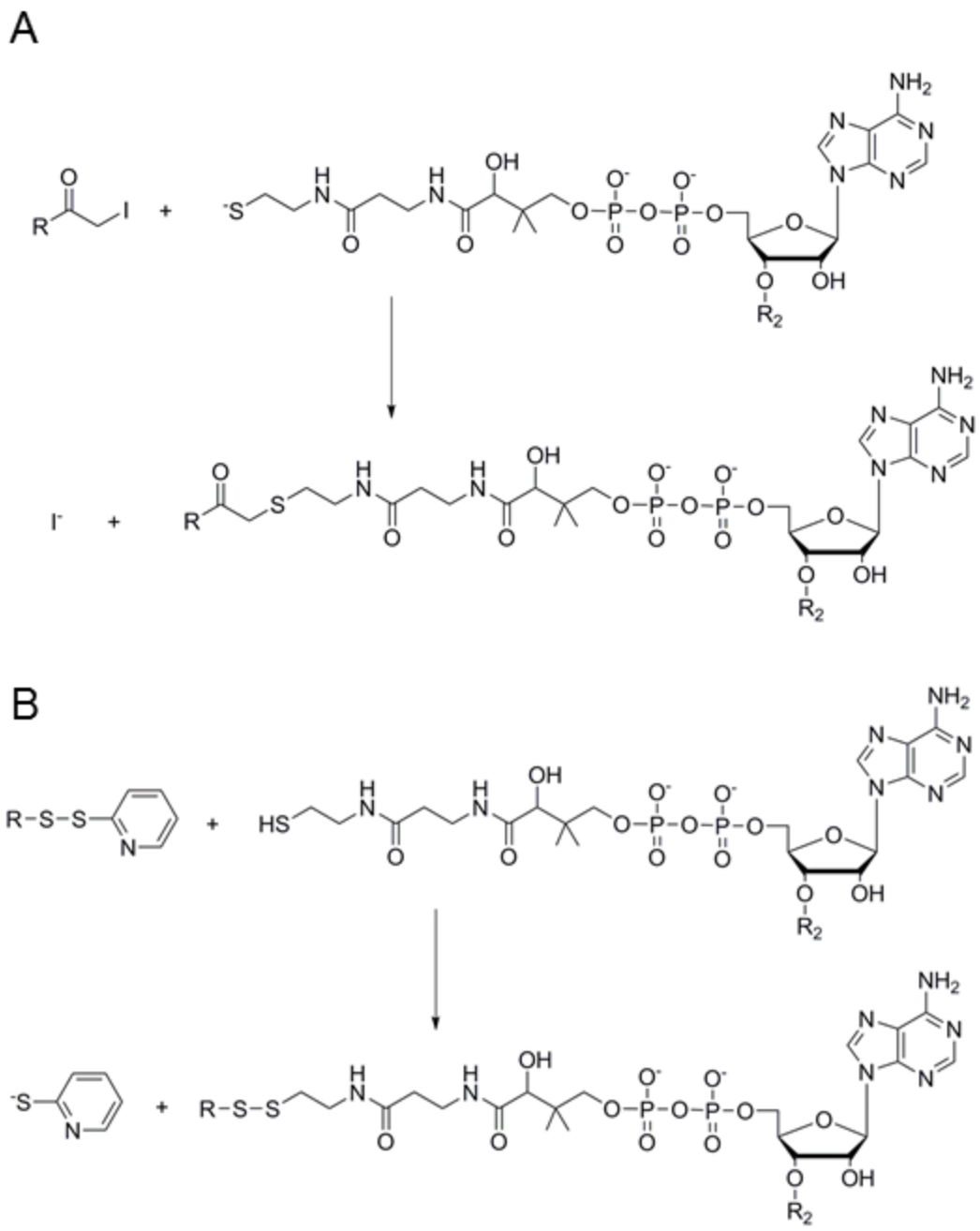


Figure 5.1 CoA-Resin Conjugation.

The thiol of CoA can be used to conjugate CoA to resins functionalized with an iodoacetyl group to form a thioether bond (A) or to form a disulfide bond if the resin is thiol functionalized (B).

In order to estimate the capacity of the beads, the iodo groups were reacted with β-mercaptoethanol. Twenty microliters of Tentagel amino resin (Chem Impex) functionalized with SIA (TGSIA, Fig. 5.2A & 5.3) were washed with excess water and then incubated at room temperature in the presence of 10 mM Tris (pH 8), and 0.006% (v/v) β-mercaptoethanol (0.8 mM) at room temperature while vortexing to monitor the release of I^- by its absorbance at 226 nm (Fig. 5.4A). Using the extinction coefficient of I^- at 14 mM^{-1} cm^{-1}, the beads were considered to contain a capacity of approximately 4 mM. The calculated capacity should be sufficient for the capture of CoA-RNA given the estimated number of molecules by Kowtoniuk et al. (2009), but it is still lower than expected as a result of possible resin derivatization inefficiency.

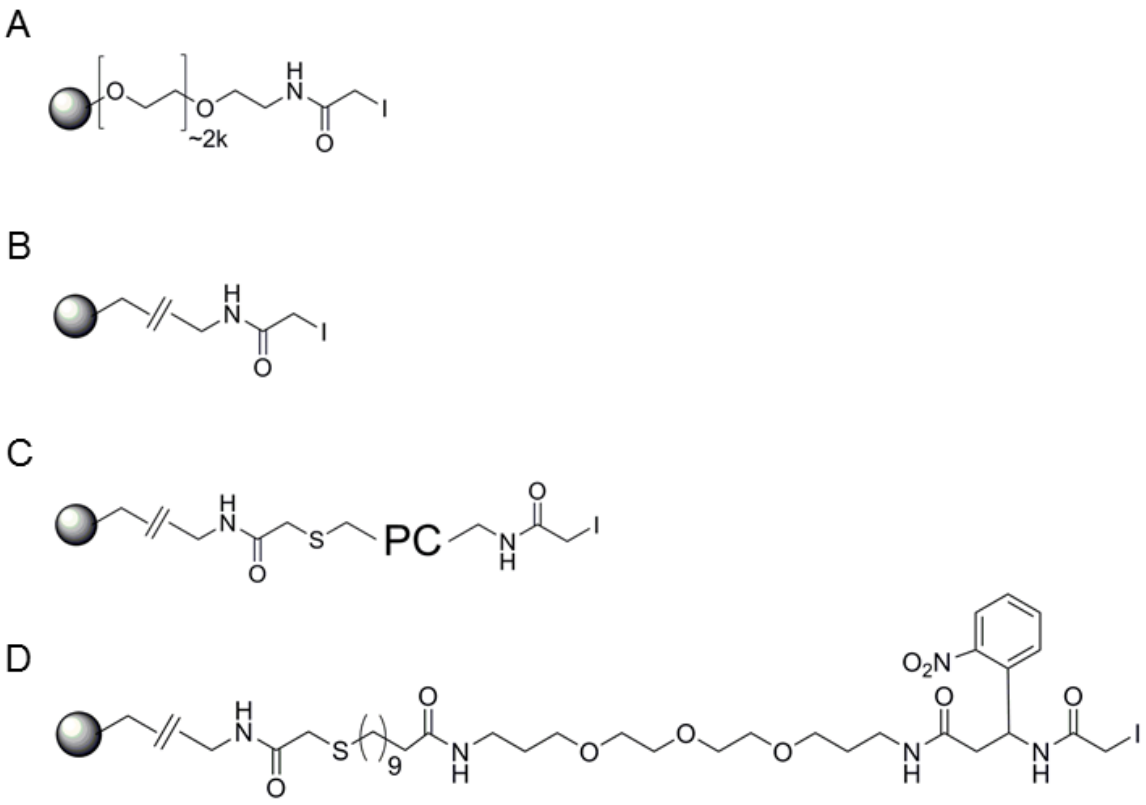


Figure 5.2 Resins for CoA-RNA Capture.

A, Tentagel-HL-SIA (TGSIA). B, UltraLink Iodoacetyl (ULIA). C, UltraLink Iodoacetyl-PC-linker-SIA (ULLI). D, ULPCIA .

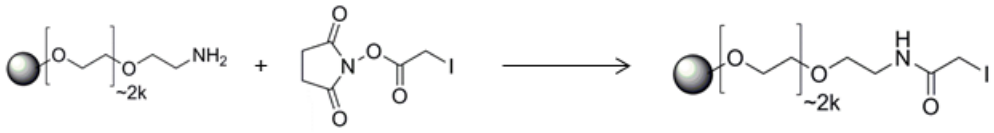


Figure 5.3 TGSIA Synthesis.

Tentagel-HL-NH2 resin combined with succinimidyl iodoacetate (SIA) to form TGSIA.

CoA-RNA was prepared by IVT in the presence of depCoA for trial capture and release from the TGSIA beads. The RNA was radiolabeled ligation of short

22mer RNA that itself had been phosphorylated by γ-^{32}P-ATP using T4 Polynucleotide Kinase (Scheme 5.2). The radiolabeled adapter will only ligate to the 3' end of the CoA-RNA, and therefore acts as an internal label (Fig. 5.5). The labeled CoA-RNA was combined with 20 µL TGSIA beads with 20 mM Tris at pH 8 in a 50 µL volume and vortexed for 30 minutes at room temperature. After CoA-RNA conjugation, the remaining iodo groups were reacted with excess β-mercaptoethanol, followed by extensive washing for multiple rounds with 40 bed volumes of water, 4 M NaCl, and 8 M urea while heating briefly at 95 °C and vortexing. The urea and salt were removed from the beads by rinsing with excess water, and the remaining radioactivity was determined to be 300 counts per minute (cpm) by liquid scintillation counter. This measurement is very low and near the detection limit of the liquid scintillation counter, indicating that only approximately 10% of the RNA was captured on the beads. The estimated yield of captured CoA-RNA presented a problem with the overall protocol. If only 10% was expected to be captured, the hopes for isolating pools of CoA-RNA from bacterial total RNA extracts may be impossible, especially when considering the low estimated amounts of CoA- and thioester-CoA-RNA in the cell. While the TGSIA did appear to capture some RNA, the estimated capacity of 4 mM was not reached by the trial with labeled CoA-RNA.

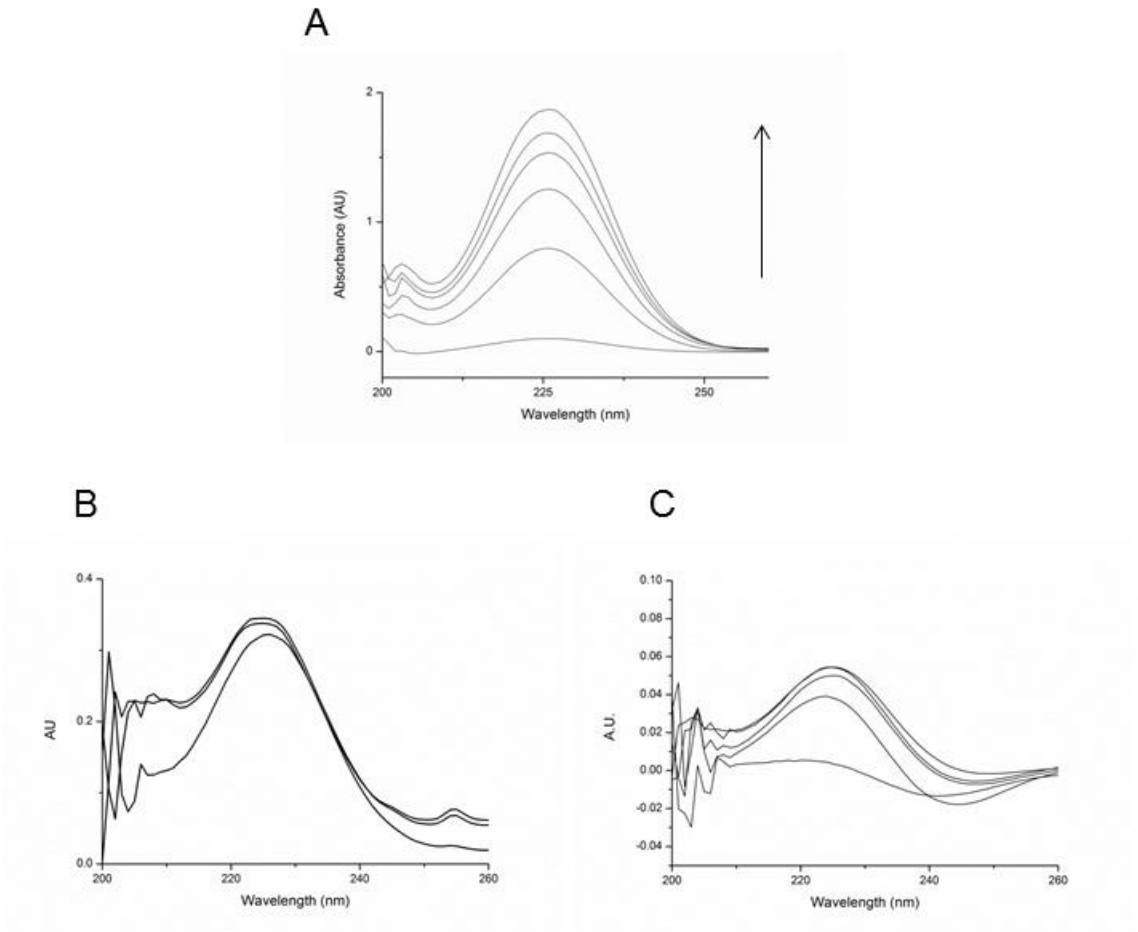
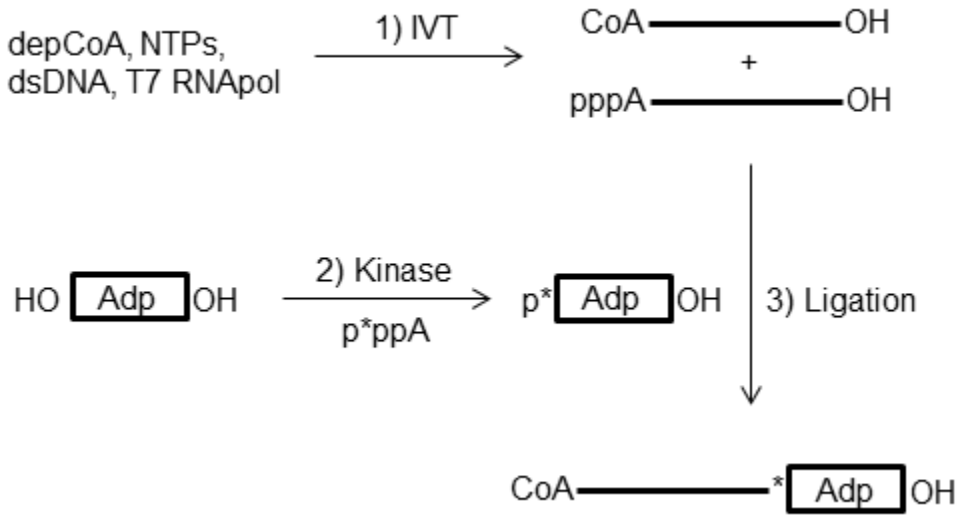


Figure 5.4 Iodo Capacity by β-Mercaptoethanol.

A, UV absorbance spectra showing increase at 226 nm indicating release of iodo from TGSIA beads over time in the presence of B-mercaptoethanol, where each line represents 5 minutes from 0 to 25 minutes total. B, UV absorbance spectra showing increase at 226 nm indicating the release of iodo from ULIA beads in the presence of B-mercaptoethanol over time, where each spectra represents 5 minutes from 0 to 10 minutes total. C, UV absorbance spectra showing increase at 226 nm indicating release of iodo from ULLI beads over time in the presence of B-mercaptoethanol, where each line represents 5 minutes from 0 to 20 minutes total.

5.2.2 Capture of CoA-RNA by ULIA

The commercially available UltraLink Iodoacetyl (ULIA) resin from Thermo was acquired to replace the TGSIA with the expectation that the capture yield would be much higher. To test for its capacity 20 µL of ULIA were washed with water and incubated in 10 mM Tris (pH 8) and 0.8 mM β-mercaptoethanol at room temperature while monitoring the absorbance at 226 nm. Using the same extinction coefficient of 14 mM^{-1} cm^{-1}, the ULIA beads were determined to have a capacity of approximately 30 mM (Fig. 5.4B). In addition to the higher capacity, the iodo was completely released within minutes of the initial contact with the higher pH buffer and β-mercaptoethanol, therefore making ULIA a much better candidate for CoA-RNA capture.



Scheme 5.2 Internal Labeling of CoA-RNA by γATP.

CoA-RNA is generated by IVT in the presence of depCoA (1). RNA oligo Adp is labeled by kinase reaction with γ32P-ATP (2) and used as donor in a ligation reaction with CoA-RNA (3).

CoA-RNA was internally labeled as previously described (Scheme 5.2) and conjugated to 100 µL ULIA beads in 20 mM Tris (pH 8) for 7 minutes at room temperature. After initial incubation, remaining iodo groups were blocked by the addition of excess β-mercaptoethanol for 7 minutes at room temperature. The beads were then washed extensively in a process similar to the TGSIA beads and the remaining radioactivity was determined by liquid scintillation counter to be approximately 30% of the input CoA-RNA. The amount of captured RNA is much higher than previously observed for the TGSIA beads in one quarter the amount of time. Considering the CoA-RNA was not purified from pppRNA generated during IVT and tends to be optimally a 50% yield mixture (Scheme 5.2), the estimated RNA captured is a promising figure for our purposes (Huang 2003a).

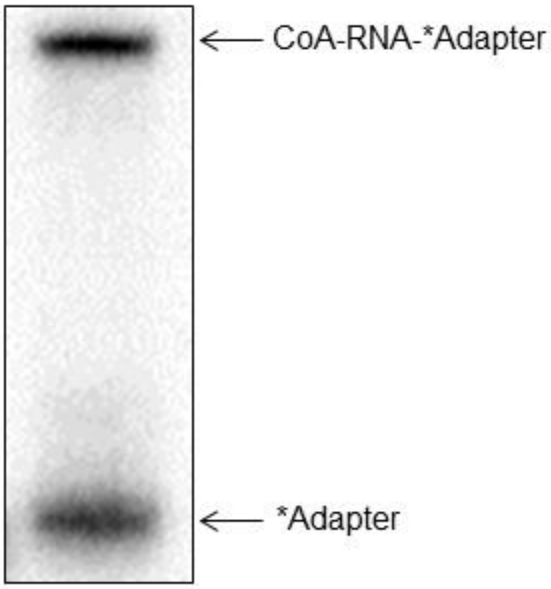


Figure 5.5 Radiolabeling of CoA-RNA by Ligation.

Autoradiograph showing the internal labeling of CoA RNA, where a short RNA adapter is labeled by T4PNK with γ-^{32}P-ATP and then ligated to the 3' end of CoA-RNA.

Despite the higher capacity for CoA-RNA capture, the ULIA beads had an inherent problem with their structure. Capture of CoA-RNA is irrelevant if the RNA cannot be released from the resin. Iodo-functionalized resin is advantageous because the formation of a stable thioether bond with CoA is a hard bond to break in comparison to disulfide bonds that can be reduced in the presence of DTT solutions. Therefore the release of CoA-RNA from iodo-functionalized resin is dependent on enzyme activity such as the murine Nudt7 enzyme described in the previous chapter. This enzyme is quite capable of hydrolyzing CoA and CoA thioesters with good efficiency, and was shown to be a

suitable candidate for our RNA release protocol. When beads conjugated with internally labeled CoA-RNA were incubated with 1X Nudix buffer and 16 μM Nudt7, however, no RNA was observed to be released as indicated by a lack of radioactivity in the reaction solution.

The ULIA beads only have a short distance between the reactive iodo group and the polyacrylamide bead itself (Fig. 5.2B), which seemed to create problems for the enzyme hydrolysis of CoA. Therefore a photocleavable linker was added to the ULIA resin to act the primary releasing of CoA (Fig. 5.2C and Scheme 5.3). Upon photocleavage, the CoA can then be hydrolyzed by Nudt7 in solution to yield the 5'-pRNA required for the subsequent processing and sequencing protocols (Scheme 5.1, Reactions 3B and 4).

5.2.3 CoA Capture by Photocleavable ULLI and ULPCIA

The ULIA resin with the added photocleavable group and a terminal iodo group (ULLI) was tested for capacity by incubating 20 μL resin with 10 mM Tris (pH 8) and 0.8 mM β-mercaptoethanol at room temperature while monitoring the absorbance at 226 nm. Using the extinction coefficient for I^- at 226 nm, the capacity was determined to be around 2.5 mM. The observed 10-fold decrease in estimated capacity of ULLI in comparison to the parent ULIA could have resulted from the manual preparation. The capacity, while low, should still be enough to capture the CoA-RNA given its small estimated concentrations by Kowtoniuk et al. (2016), and the release of the iodide was noted to occur quickly within 10 minutes of incubation (Fig. 5.4C). Considering the release of the target molecule is as important as the high capacity for capture, the beads needed to be tested

for their photocleavage efficiency. Aliquots of 10 µL of different iterations of ULIA derivatized with photocleavable linkers, ULLI and ULPCIA (Fig. 5.2C and D), were conjugated with 1 mM CoA in the presence of 20 mM Tris (pH 8) at room temperature, vortexing for 5 minutes by hand immediately upon addition of the RNA/Tris solution and then another hour afterwards. After conjugation, the absorbance at 260 nm was tested using the extinction coefficient of adenosine 15 mM^{-1} cm^{-1} to determine the amount of CoA not conjugated to the beads. Upon washing, unconjugated CoA continued to be quantified in addition to released iodide determined by the absorbance at 226 nm in order to quantify the capacity of the beads. Capacity was calculated by adding the difference between the initial CoA and CoA that was washed out added to the amount of iodide released in solution. The ULLI beads were determined to have a 14 mM capacity with 12 mM CoA estimated to be conjugated, while the ULPCIA beads were determined to have a capacity of 23 mM with 15 mM CoA estimated to be conjugated.

Each resin was split into two aliquots, where one would be treated with 16 µM Nudt7 in 1X Nudix buffer for 30 minutes at 37 °C, and the other exposed to a 365 nm UV lamp for 5 minutes on ice. The resulting solutions were tested for their absorbance at 260 nm to detect released CoA, and the ULPCIA, but not the ULLI beads, were noted to have released CoA upon exposure to UV light (Fig. 5.6A). The released fragments were analyzed by electrospray ionization mass spectrometry (ESI-MS) where the expected photocleavage product of 824.14 Da was observed as a peak with m/z of 823.09 representing the $[M\text{-}H^+]$ species, as well as a peak with m/z of 411.16 representing the $[M\text{-}2H^+]/2$ species (Fig. 5.6C).

Photocleavage reactions were quantified by the absorbance of released CoA at 260 nm, and it was determined that approximately 1.5 mM CoA was released after 5 minutes of UV exposure, and another 1.4 mM released after 10 minutes of exposure. Based on the initial 1 mM CoA concentration in the reaction volume of 120 µL and the bead volume of 5 µL, the capture of 100% of the available CoA would result in a 24 mM concentration on the beads. Given the estimated amount of conjugated CoA to be 15 mM, and the combined 2.4 mM CoA observed to be released upon photocleavage, the ULPCIA beads were determined to release approximately 16% of the total possible CoA.

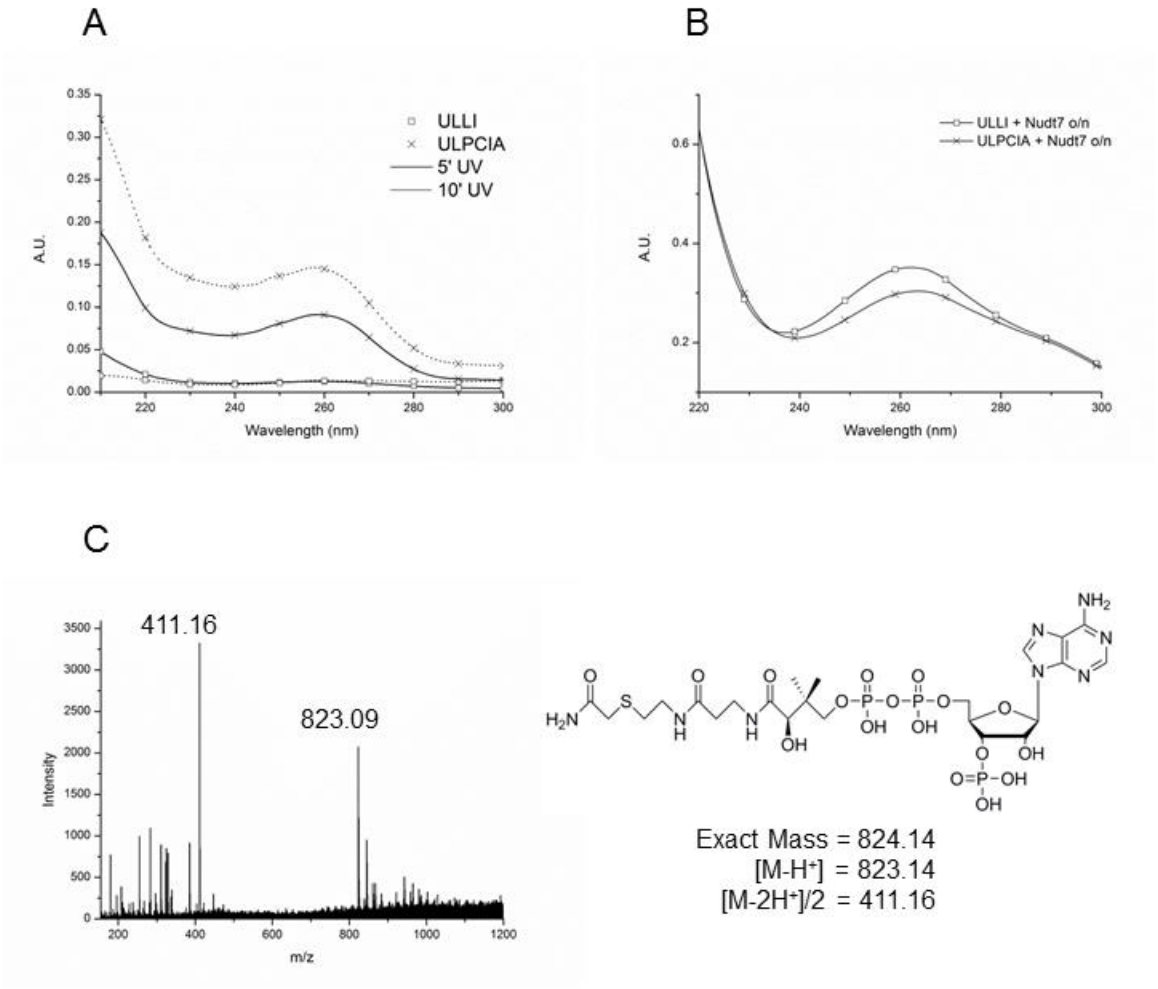
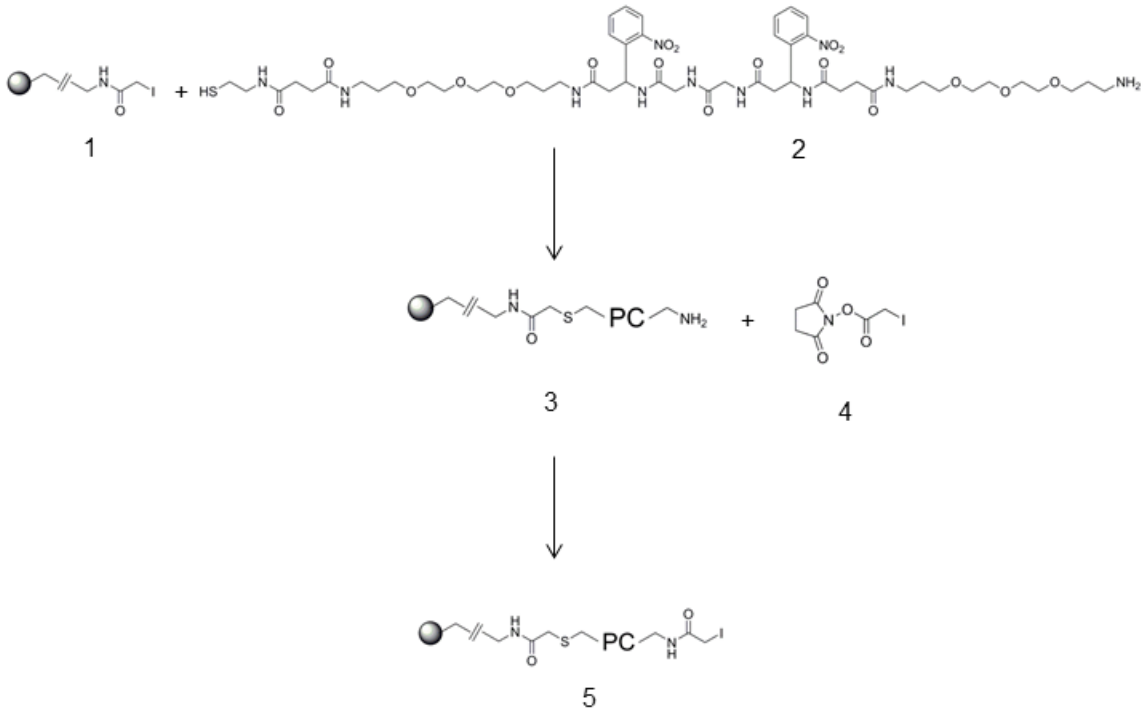


Figure 5.6 CoA Release from ULLI and ULPCIA.

A, UV absorbance spectra of CoA released by photocleavage from ULLI beads (square boxes) and ULPCIA beads (x) after 5 minutes (solid line) and 10 minutes (dashed line). B, UV absorbance spectra of 3', 5'-ADP released by overnight treatment of ULLI (box) and ULPCIA (x) by Nudt7. C, ESI-MS of photocleavage product from ULPCIA beads.

The method of photocleavage was apparently unsuccessful for the ULLI beads altogether, for no observable CoA appeared in the reaction solution after photocleavage (Fig. 5.6A). The reasoning for the low photocleavage could be

due to the inability for the photons to penetrate the polyacrylamide resin completely, and therefore only the plane exposed to the UV light was capable of reacting. The aliquot of CoA-conjugated ULLI and ULPCIA beads that were subjected to Nudt7 digestion were also analyzed by UV absorbance at 260 nm for the 3', 5'-ADP released, and it was determined that 3 mM ADP was released from the ULLI beads, and no detectable ADP was released by Nudt7 hydrolysis from the ULPCIA beads. The two resins were incubated overnight in the same Nudt7 reaction conditions at 37 °C, and subsequently another 3.8 mM ADP was seen to be released from the ULLI beads for a total of 6.8 mM, and 3.2 mM ADP was released from the ULPCIA beads (Fig. 5.6B). Given the estimates for CoA conjugated to each resin, Nudt7 was able to hydrolyze 56% of the CoA from the ULLI beads in 16 hours at 37 °C, and 12.5% of the CoA from the ULPCIA beads in the same amount of time (Table 5.1).



Scheme 5.3 ULLI Synthesis.

The UltraLink Iodoacetyl resin (1) was reacted with the PC-linker (2) with two nitrophenyl groups flanked by units of 4,7,10-Trioxa-1,13-tridecanediamine (TOTDDA), with one of the two TOTDDA groups conjugated to a succinate and cystamine to leave a free thiol that can react with the iodo from (1) forming a thioether linkage. The resulting molecule (3) was then combined with succinimidyl iodoacetate (SIA, 4) to leave a free iodoacetyl group at the terminus opposite the resin (5).

Table 5.1

CoA Released from Iodo-functionalized Resins

Resin	Release Method	Time	% Released
ULIA	Nudt7	30 min	-
ULLI	Nudt7	30 min	25%
		Overnight	56%
ULPCIA	Nudt7	30 min	-
		Overnight	12.5%
ULLI	UV	10 min	-
ULPCIA	UV	10 min	16%

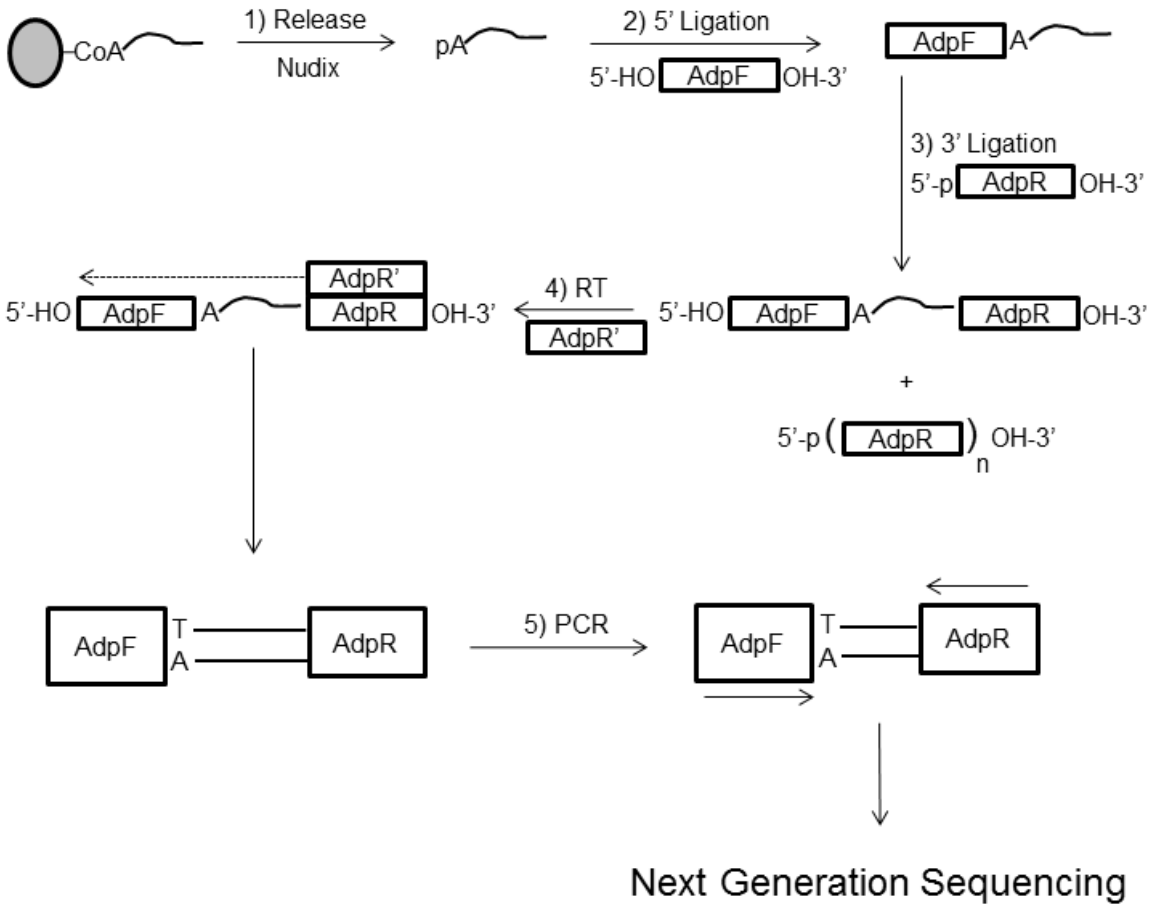
The amount of CoA conjugated to each resin was determined to be less than the capacity. The most efficient resin for capturing CoA was the ULLI resin, in which approximately 86% of the available capacity was actually conjugated with CoA. Unfortunately, the photocleavage from the ULLI resin failed, and Nudt7 could only hydrolyze 56% of the conjugated CoA when incubated for an extended amount of time. The incubation period for Nudt7 based on previously characterized enzyme kinetics infer that hydrolysis should be completed at these concentrations in under 30 minutes (Gasmi and McLennan 2001). Therefore based on the fact that 30 minutes and 16 hours incubation yielded the same

amount of hydrolyzed CoA, it is likely that the enzyme cleaved the maximum possible CoA in the first few minutes of the reaction with the beads, and the remaining time spent in incubation yielded no results. During the incubation of the enzyme with the beads, the substrate is not in solution but rather in suspension where the beads themselves settle over time and can exclude the enzyme from contact with its substrate. Continuous agitation of the resin should be carried out in order to keep it in suspension and maximize the contact of the enzyme with the beads, thereby increasing the yield of hydrolysis. The ULPCIA beads, on the other hand, proved to be a poor resin for use with Nudt7 hydrolysis. Considering the design of the resin was not for Nudt7 treatment but for photocleavage, this discrepancy is understandable. The lack of Nudt7 hydrolysis from this resin may come from the extended hydrophobic region of the photocleavable linker, which may have interfered with enzyme binding of the substrate, or caused non-covalent interaction with the enzyme preventing its activity.

5.2.4 Ligation Optimization

The CoA-RNA processing protocol was prepared separately from the capture protocol for optimization purposes. Because of the amount of steps involved in the entire process, it was considered unfeasible to begin with bead conjugation and analyze efficiency at the final step of PCR. The process of optimizing CoA-RNA ligation would be undertaken first. The principle of adapter ligation to CoA-RNA is to add priming sites that can later be amplified by PCR after reverse transcription (Scheme 5.4). The priming sequences must be

distinctly different from the sequence of the CoA-RNA or else the PCR will be problematic. Considering that the sequences are wholly unknown, there is much uncertainty in the design of adapters. The CoA-RNA would ideally be ligated on its 3’ end by a phosphorylated adapter that contains a 5’-phosphate, and itself is blocked from self-ligation by either sodium periodate oxidation of the diol (Scheme 5.5) or by amine derivatization which can be arranged when ordering synthetic RNA. After the 3’ adapter ligation, reverse transcription can be carried out using a primer complementary to the 3’ adapter. The resulting cDNA could then be further ligated to another 3’ adapter, considering the 3’ end of the cDNA is the complement to the 5’ end of the CoA-RNA. Following the second ligation, PCR would be used to amplify the cDNA library, and the resulting products can be submitted for NGS library preparation.

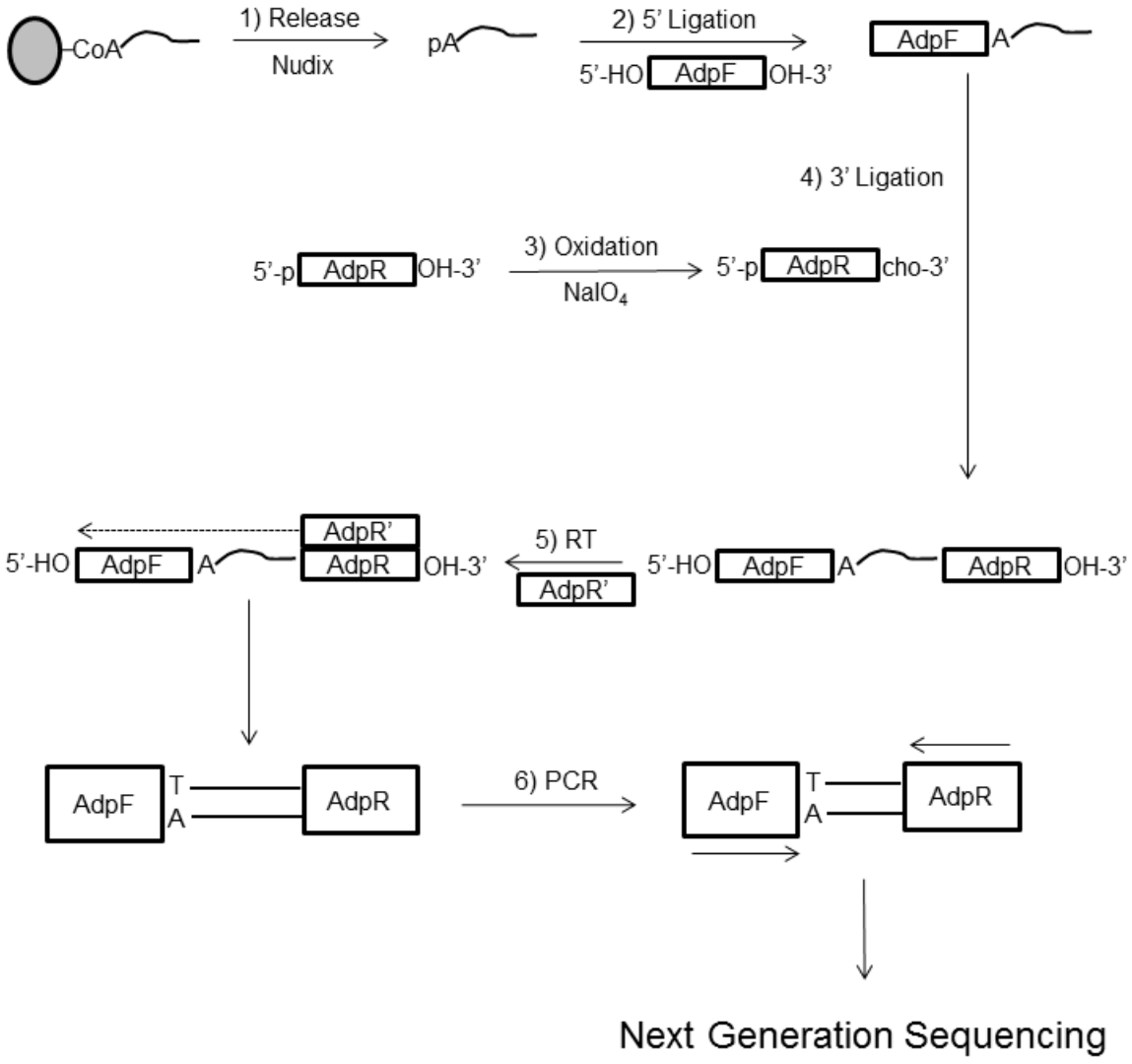


Scheme 5.4 Capture Protocol Using Double Ligation.

Capture protocol using double ligation. RNA captured by resin and released by Nudix hydrolysis (1) leaves a 5'-pRNA capable of 5' ligation (2) to a 5'HO-adapter. The subsequent ligation of the 3' adapter (3) is problematic due to the possible multimer formation by the 5'-pRNA adapter used without blocking its 3' end. RT (4) after ligation can be used to generate cDNA that can then be amplified by PCR (5) to generate the NGS library.

Adapter ligation is key when capturing and amplifying RNA of unknown sequences. A typical mRNA identification protocol would involve RT using a poly(dT) primer that would prime the poly(A) tail of mRNA. Considering the lack of knowledge of these CoA-RNA sequences and the probability for their lack of a poly(A) tail, adapters with high ligation efficiency should be used that will serve

as priming sites for RT. Two RNAs were generated to test the efficiency of periodate treatment of RNA that serves to block RNAs from multimerization via intermolecular self-ligation (Scheme 5.5). A 40 nt RNA (40mP) was generated by IVT in the presence of AMP in a 4 molar excess to ATP. The increase of AMP in the reaction mixture should result in a 4:1 ratio of 5'pRNA to 5'pppRNA. Similarly, an 80 nt RNA (80ppp) was generated using standard IVT conditions to produce a homogeneous stock of 5'pppRNA. After purification of both by 8% denaturing PAGE, 40mP was treated with 10 mM $NaIO_4$ for 40 minutes on ice in the dark to oxidize the *cis*-diol of the 3' terminal ribose into a dialdehyde, and the reaction was quenched by ethanol precipitation. The resulting RNA p40cho was combined in equimolar amounts of 80ppp in a ligation reaction at 37 °C for 50 minutes and the reaction was visualized by 8% denaturing PAGE (Fig. 5.7A). The success of the periodate treatment is noted by the single band resulting from ligation that represents the 120 nt RNA. In this model, the 80ppp RNA represents our CoA-RNA which is incapable of 5' ligation, and the p40cho represents our adapter to be ligated to the 3' end of the CoA-RNA.



Scheme 5.5 Capture Protocol Using RNA Oxidation and Double Ligation.

Capture protocol using RNA oxidation and double ligation. Similar to the previous schemes, where captured RNA is released by Nudix digestion (1) followed by ligation of the 5' adaptor (2). Here, the 3' adaptor is subjected to oxidation by periodate to generate a dialdehyde on the 3'end of the 3'-adaptor (3) to prevent self-ligation and multimerization upon 3'-adapter ligation (4). Removing such multimers allows the cleaner success of RT (5) and PCR (6) to generate the NGS library.

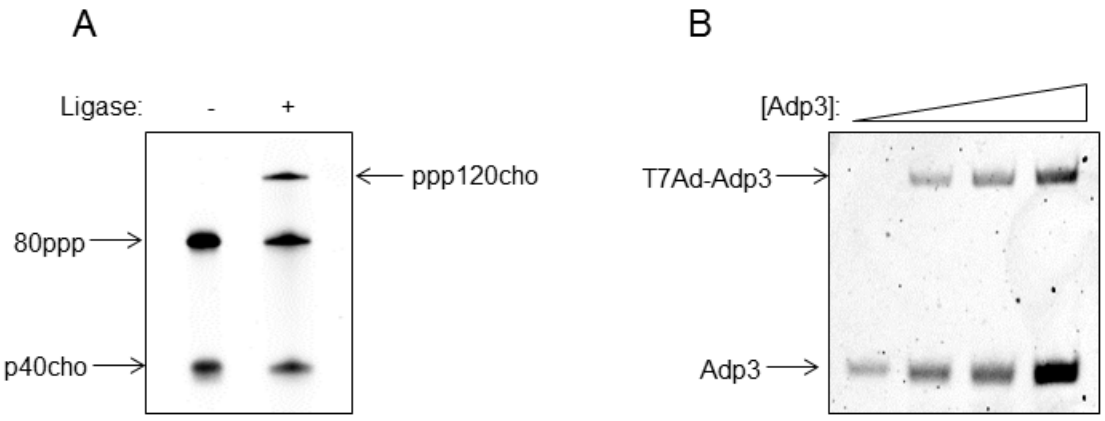


Figure 5.7 Adaptor Ligation.

A, Ligation of periodate-treated 5'-monophosphate RNA (p40cho) and 5'-triphosphate RNA (80ppp). B, Ligation of adenylated pDNA adapter Adp3 at increasing concentrations with in vitro transcribed adapter T7Ad.

The ligation of the 3' adapter in some miRNA capturing protocols (rapid amplification of cDNA ends, or RACE) involves the use of a DNA adapter, which is still capable of ligation to a ssRNA molecule at a lower efficiency. The differences in efficiency of reactions in solution and on solid surfaces such as resin can be attributed to the diffusion barriers that prevent the interaction of the substrates and enzyme. In the case of ligation, the ligase enzyme must come in contact with the donor molecule (the adapter), the acceptor molecule (the CoA-RNA), and ATP in order for the reaction to be catalyzed. Using an adenylated adapter is a way of increasing the efficiency by relieving the reaction's need for ATP. A short 20 nt 5'-pDNA oligo for this purpose (Adp3) was obtained from IDT modified with a 3'-amino modification to block multimerization during ligation. The DNA was incubated with 30 mM $MgCl_2$ and 120 mM adenosine phosphoimidazolide (ImpA) at 50 °C for 3 hours and purified by an Amicon 3K

spin column followed by ethanol precipitation. The resulting product was resolved on 10% denaturing PAGE from which it was purified and combined with an IVT-produced adapter (T7Ad) in a ligation reaction at 37 °C for 1 hour (Fig 5.7B).

Increasing amounts of the adenylated pDNA Adp3 increased the single ligation product with T7Ad in the absence of ATP as shown in Figure 5.7B, therefore indicating the adenylation of Adp3 had been successful. Despite the success of ligation with the adenylated DNA oligo Adp3 with ligation, further experiments were carried out to determine which of a group of different short synthetic RNAs would give the best ligation efficiency. Ten different samples of synthetic RNA were treated with T4PNK to add a 5'-phosphate group to become ligation donors, and were subjected to ligation reactions to determine reaction efficiency based on the degree of intramolecular ligation that occurred. Based on the sequences of the RNA and the observed ligation efficiencies, the RNA GL1a was determined to be the best suitable adaptor.

5.2.5 Reverse Transcription Optimization

The process of RT-PCR would require some optimization when considering its success was coupled to the success of adapter ligation. To first test the efficiency of RT, a known RNA sample that is an 800 bp fragment of EGFP mRNA was generated via IVT in the presence of depCoA to generate CoA-EGFP and purified from the reaction mixture using an M100 spin column. The CoA-RNA was ligated with adenylated Adp3 and purified by an M100 spin column, treated with 16 μM Nudt7 for 1 hour at 37 °C, and immediately ligated to T7Ad for 30 minutes at 37 °C. A parallel reaction was carried out simultaneously

without Nudt7 reaction to be used as a negative control. Both RNA samples were subjected to RT using primers complementary to the 3’ adapter Adp3, followed by PCR on the resulting cDNA with primers complementary to both adapters. The resulting agarose gel indicated that while PCR had been successful, the 800 bp target band was not amplified, but instead a smaller band below 150 bp had appeared in both experimental samples and control samples where RT was not carried out (Fig. 5.8A).

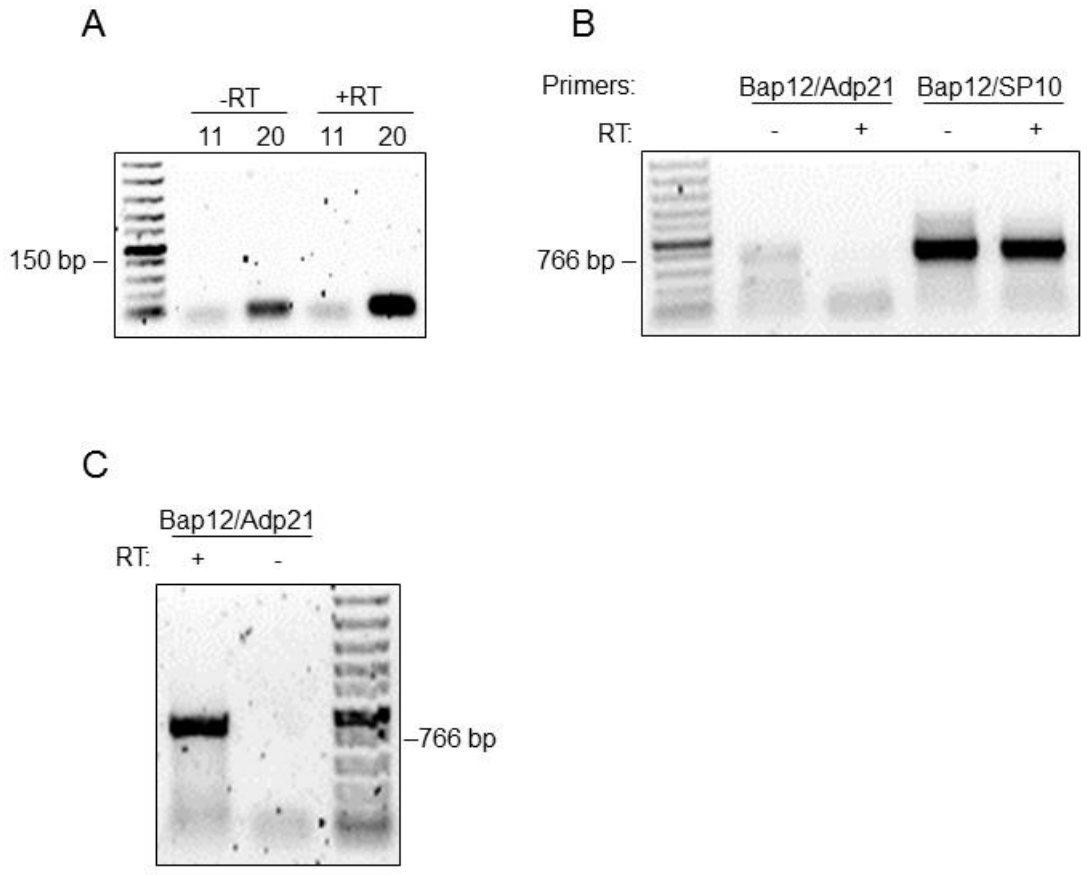


Figure 5.8 RT-PCR of Adapter-Ligated RNA.

A, Results of PCR on CoA-EGFP-RNA sample after RT along with controls (-RT). B, Results of PCR on CoA-EGFP-RNA sample with and without RT using primers specific for the sequence itself (Bap12/SP10) or primers requiring adapter ligation for success (Bap12/Adp21). C, Results of PCR on CoA-EGFP-RNA sample after with or without RT treatment after the removal of the dsDNA template for IVT by restriction digestion.

The PCR was repeated on obtained cDNA with different primers that were complementary to the EGFP sequence (Bap12 as forward primer, SP10 as reverse primer), as well as one primer complementary to the EGFP sequence (Bap12 forward) and another complementary to the adapter (Adp21 reverse), and the resulting agarose gel indicated a band at the expected size of

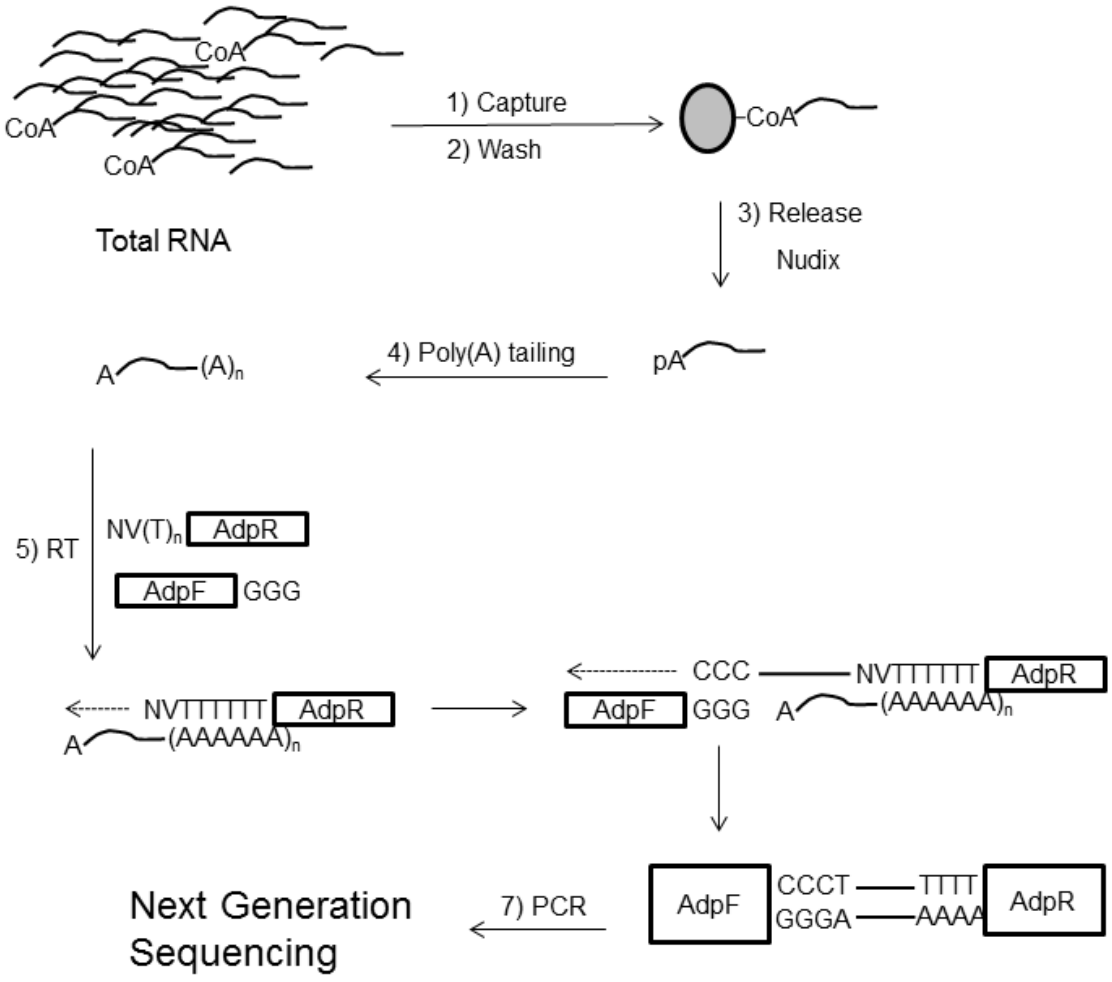
approximately 900 bp (Fig. 5.8B). Given the appearance of the band in a negative control that did not undergo RT and the additional lack of amplification using the Adp21 primer, it seemed the RT did not succeed due to the failure of the initial ligation reaction. The bands observed in Figure 5.8B likely resulted from the presence of the dsDNA template for IVT that was never purified from the reaction mixture for this testing. RNA generated by IVT is typically purified by 8% denaturing PAGE that can resolve down to a single nucleotide, but the CoA-EGFP RNA was so large that this method became unfeasible. Therefore the M100 spin column used instead to purify the RNA from the buffer and enzyme was not enough to discriminate against the dsDNA template that is larger due to the incorporation of the promoter sequence.

The process was repeated excluding Nudt7 after the double digestion of CoA-EGFP RNA with AluI and HaeIII restriction enzymes to remove the dsDNA template from the RNA solution. After ligation of adenylated Adp3 and RT using its complementary primer Adp21, PCR with a specific forward primer Bap12 with the reverse primer Adp21 was carried out to produce the expected 800 bp EGFP band (Fig. 5.8C). The lack of a band in with from control samples that did not undergo RT in addition to the band produced using Adp21 as a reverse primer indicated that the 3’ adapter ligation was successful, as was the subsequent RT (Fig. 5.8C). While the 3’ adapter was shown to be ligated when the process was repeated, the secondary ligation of the 5’ adapter remained problematic. This could have been due to the requirement for Nudt7 digestion to be complete before the ligation could occur. Nudt7 digestion in solution is a quick reaction with

low concentrations of substrate (< 1 mM), and therefore should not be the issue. The secondary ligation reaction did take place directly following Nudt7 digestion, and both were carried out in the same buffer. Nudt7 activity was testing in the ligation buffer and CoA hydrolysis was not observed to be impeded, and Nudt7 was also demonstrated to have no activity against ATP, thereby again the Nudt7 hydrolysis of CoA was not likely the problem. It seemed more feasible that the problem was the ligation itself, as it has been known to occur at a low efficiency depending on the sequences being used (Munafo and Robb 2010). Such reasons warranted earlier studies to optimize ligation conditions to maximize the efficiency before moving forward. However considering that the CoA-RNA sequences are unknown, the 3' adapter ligation can never truly be optimized due because of the variable of the 3' terminal nucleotide. On the other hand, the 5' ligation should be easier to predict due to the fact that the donor will always be adenosine. Unfortunately, previous studies have shown A to be a poor donor in T4 RNA Ligase reactions, with only ~16% of the activity as cytosine (Romaniuk et al. 1982).

The ligation of two adapters to the CoA-RNA following Nudt7 digestion was determined to be a problematic approach to the capture protocol. Despite Nudt7 maintaining activity in a variety of buffers, the ligation of the 5' adapter to the RNA was never successful based on the lack of results from RT-PCR using primers complementary to both adapters. For this purpose, the template-switch method of RT was employed (Scheme 5.6). As stated previously, known sequences flanking the unknown CoA-RNA sequences are required for the

success of PCR amplification of the cDNA resulting from RT. While these sequences can be added to the target RNA through ligation, the template-switch (TS-RT) mechanism of reverse transcriptase can be used to add a known sequence to the 5’ end by itself. Once reverse transcription is primed and the enzyme reaches the end of the RNA template, it extends the cDNA with 3 cytosine residues (Ko and Lee 2006). Carrying out the RT reaction in the presence of an addition DNA oligo containing 3 guanosine residues at its 3’ end allows this oligo to anneal to newly generated cDNA with terminal cytosines, and therefore the enzyme continues to generate the single cDNA by switching templates from the original RNA to the secondary adaptor known as the TS oligo Adp26.



Scheme 5.6 Capture Protocol Using Template Switch-RT.

After the capture of CoA-RNA (1) and extensive washing of nonspecifically interacting RNA from the resin (2), the CoA-RNA is released from the beads by Nudix digestion (3). The following molecule is adenylated by poly(A) polymerase (4) and subjected template-switch RT (5) to add the adapter sequences for PCR (7) and preparation of the NGS library.

In addition to the TS-RT protocol, a ligation-free protocol for the addition of the 3' adapter was adopted. Using the highly efficient enzyme poly(a) polymerase, the captured RNA can be tailed with multiple adenosines, and then a reverse primer to anneal to this tail can be used to initiate RT. Because this does not satisfy the installment of a known priming sequence for downstream

PCR amplification, a primer containing a poly(dT) region downstream of a known sequence can be used so that the subsequent cDNA contains a priming region followed by a patch of thymines for PCR amplification of the unknown CoA-RNA sequence (Scheme 5.6). Due to the inherent problems with amplifying DNA homopolymers with DNA polymerase, the poly(dT) region of the primer will be limited to a lengths under 30 nt by design, and be followed by the sequence VN, where V is any nucleotide besides T, and N is any nucleotide (known as Adp13). The Adp13 primer structure is therefore a known priming sequence for later PCR followed by 12 thymines and VN, allowing the annealing of the primer to the interface of the original RNA sequence and the additional poly(A) tail and the addition of an adapter sequence to the resulting cDNA. The exclusion of ligation in the RNA processing protocol was therefore expected to increase its overall success considering ligation is a problematic and inefficient process.

A 150 nt RNA was transcribed via IVT in the presence of depCoA and purified by an M30 column to produce 150 nt CoA-RNA. After restriction enzyme treatment to remove the DNA template, the CoA-RNA was subjected to polyadenylation by poly(a) polymerase and RT using Superscript IV Reverse Transcriptase (Thermo) in the presence a reverse poly(dT) primer Adp16 and the TS-primer Adp26 followed by PCR amplification using primers PCF1 and PCF2 that are complementary to the adapter sequences present in Adp16 and Adp26, respectively. Early experiments using the template switch RT protocol with MMLV (Promega) yielded a mixture of bands, due to nonspecific interactions of the TS-primer with the template as noted by other investigators (Matz et al.

1999). The appearance of off-target products is a problem, considering that they may amplify more efficiently than true captured cDNA, and there is no way of distinguishing these artifacts from the captured CoA-RNA until after the samples have been sequenced.

The temperature of RT needed to be increased in order to reduce the nonspecific activity of the TS oligo, and for this the more thermostable enzyme Superscript IV Reverse Transcriptase (SSIV) was obtained from Thermo. General RT protocols dictate that denaturing of secondary structure and annealing of primer to template need to occur outside of the presence of the RT enzyme and the buffer which contains Zn^{2+} at a high pH. If these are put in the presence of the captured RNA for the 10 minute denaturation step at 70 °C, the RNA will likely be hydrolyzed. As such, after the denaturation, the samples are generally put on ice to cool before the addition of the buffer and enzyme, and afterwards RT proceeds. When the efficiency of the RT protocol was tested by varying the extension temperature from 46 °C to 70 °C as well as the method of enzyme addition, the ideal conditions that minimize off-target products while maximizing target cDNA yield was determined to be direct addition of the enzyme following RNA denaturation and annealing followed by incubation at 60 °C (Fig. 5.9). As temperature of incubation increases, the amount of off-target (noted as a smear between 75 and 150 bp) decreases. The amount of off-target product continues to decrease as the incubation temperature rises above 60 °C, however the total yield of cDNA tends to decrease as well. The direct addition of enzyme without cooling the reaction mixture on ice also clearly shows a better total yield

in addition to a more favorable target to off-target product ratio. Therefore the TS-RT protocol was determined to be optimal at 60 °C with the direct addition of the enzyme.

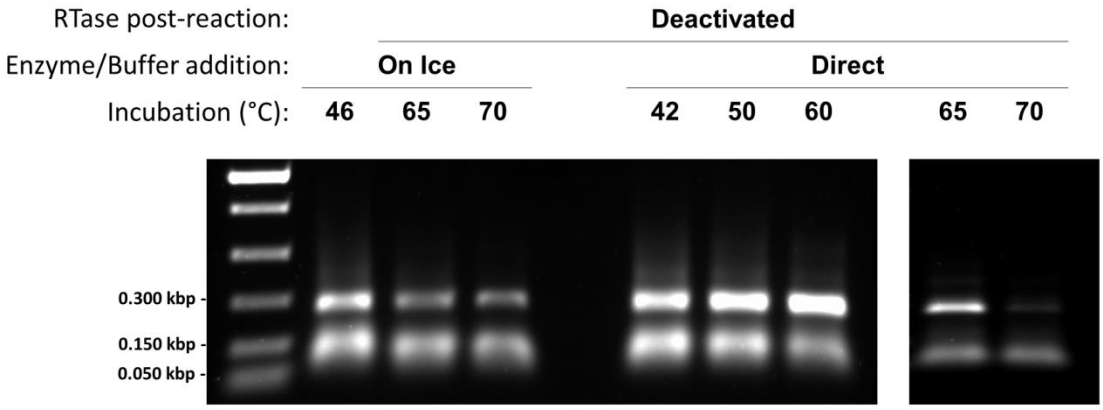


Figure 5.9 Template Switch-RT Optimization.

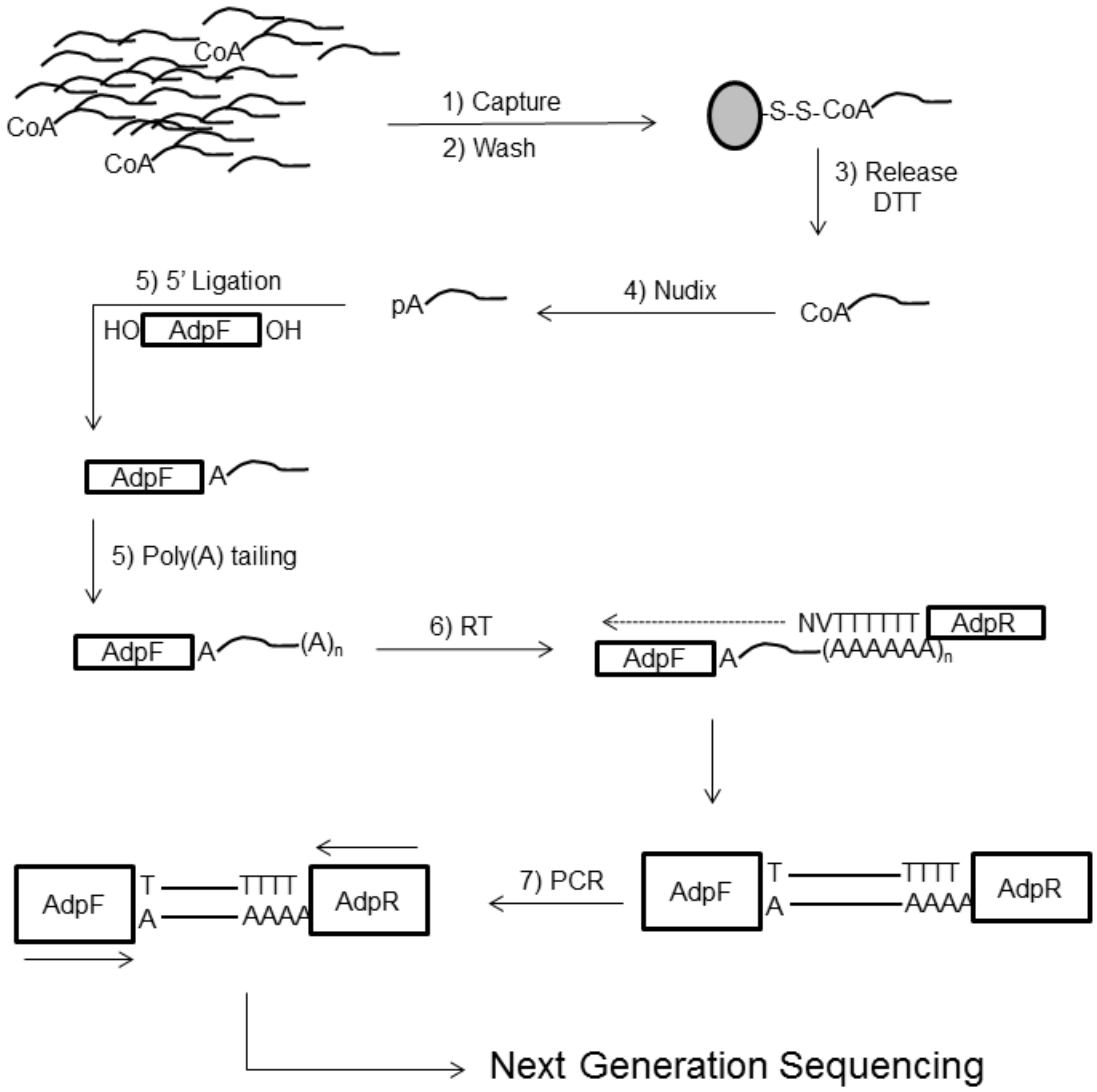
Different parameters of template-switch RT (TS-RT) were tested in order to optimize the total target cDNA generated as well as increase the ratio of target to off-target by-products. Samples were incubated at various temperatures during the RT reaction and were composed by the addition of enzyme and buffer following denaturation either after cooling on ice or by direct addition. The expected product of 300 bp and the off-target products between 75 and 150 bp were visualized after 30 cycles of PCR following RT.

Based on the troubles noted in the specific capture protocol of Cahova et al. (2009), the observed artifacts generated by the use of template-switch RT were problematic. The issue of purity up to this point has been of paramount importance when optimizing capture and release protocols. The purity during RNA processing and generation of the NGS library is something that should not be compromised, for these small sized off-target PCR products can quickly overtake the actual targeted library during PCR, and the subsequent NGS results can be tainted. While the NGS protocols do offer quite a depth of sequencing on

the order of tens of millions of reads, the submission of high quality NGS libraries will result in high quality sequencing results. This should be the overarching purpose of the CoA-RNA protocol, and it should not be compromised by DNA amplification techniques, of all things.

5.2.6 E. coli CoA-RNA Capture

Once all of the components of the CoA-capture and RNA processing protocol were optimized individually, the template-switch RT protocol was decided against. While convenient due to the exclusion of low-efficiency ligation reactions, the additional artifacts resulting from PCR cannot be tolerated if it is possible to avoid them. In light of the previously stated determination for the utmost purity and quality of the CoA-RNA sample, the previously assessed resins were decided against as well, for they could only at the highest estimate facilitate 56% release of the CoA-RNA with additional enzyme treatments. Instead, the thiol sepharose 6B resin was decided upon for the CoA-RNA capture (Scheme 5.7). While incomparably simpler than any of the other resins tested, it has been routinely used in the past to purify CoA-RNA from a heterogeneous pool of RNA from IVT protocols for SELEX as well as other purposes (Coleman and Huang 2005, Coleman et al. 2005, Coleman and Huang 2002).

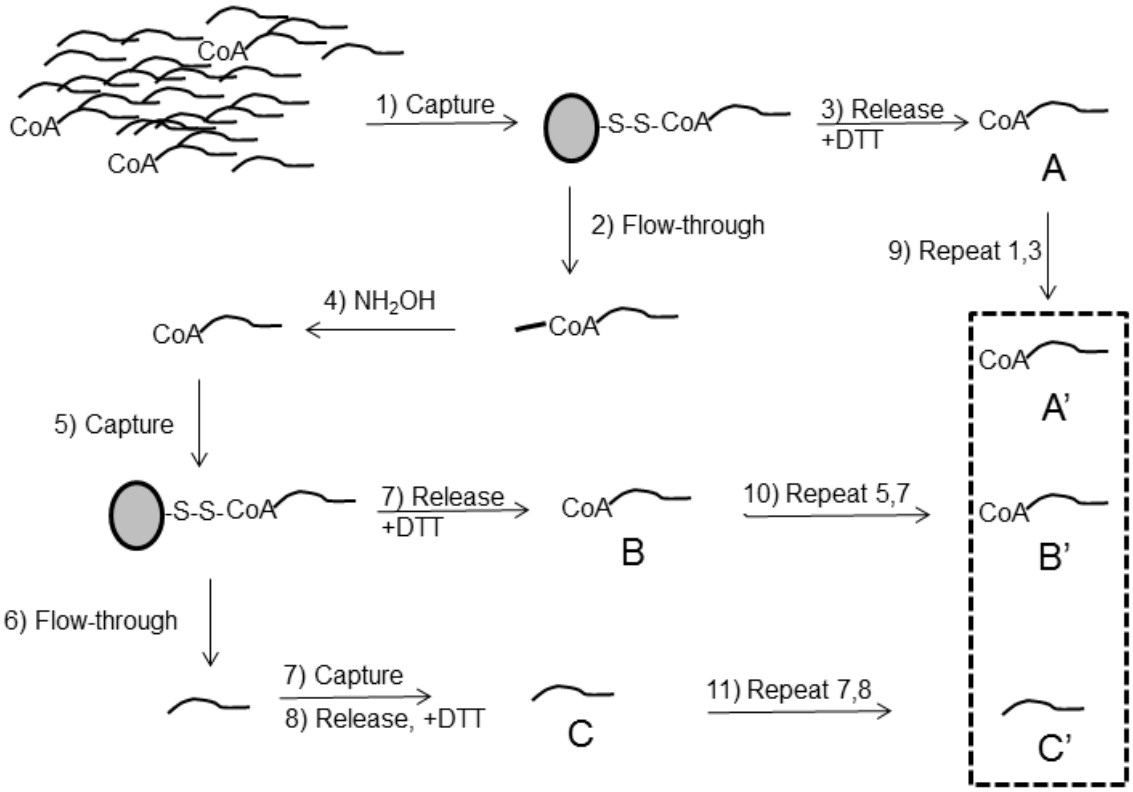


Scheme 5.7 Capture Protocol Using Single Ligation and Thermostable RTase.

The CoA-RNA is captured by disulfide bond formation (1) to the resin, and the nonspecifically interacting RNA is washed away (2). Upon the release of the RNA by the addition of reductant (3), the CoA-RNA is converted to 5'pRNA by Nudix hydrolysis (4) so that the 5' adaptor can be ligated (5). A poly(A) tail is added (5) to act as a seeding region for the RT primer AdpR in order to add the reverse priming sequence (6) so that the resulting cDNA can be amplified by PCR (7) to generate the NGS library.

Total RNA isolated from *E. coli* High-Control BL21(DE3) cells was purified by thiol sepharose resin to capture the free thiol group with a disulfide bond

(Scheme 5.8). The washing of the column with DTT after capturing to elute is a characteristically simpler method of thiol purification, and has previously been used to purify CoA-RNA from triphosphate RNA made through IVT (Huang 2003a). The isolated total RNA from *E. coli* was bound to thiol sepharose (Column A) in the presence of 70 mM Tris (pH 8) while collecting the flow-through. The column was washed extensively to remove nonspecifically bound RNA with excess 3 M NaCl, 4 M sodium acetate (NaAc, pH 5.2), 4 M urea, and water, followed by three rounds of excess 3 M NaCl and water, and then 3 M NaCl, 4 M NaAc pH 5.2, 4 M Urea, and water. After draining the column, 0.5 mL of 100 mM DTT in 100 mM Tris (pH 8) with 10 µg glycogen as a carrier was added to the column to release the CoA-RNA. The column was washed with the addition of 400 µL 0.4 M NaAc (pH 5.2) and 10 µg glycogen, and the collected flow-through was precipitated by the addition of ethanol. The initial flow-through collected was devoid of CoA-RNA, but still contained some thioester-CoA-RNA. Therefore the initial flow-through was treated with 50 mM hydroxylamine and 20 mM DTT for 15 minutes at room temperature to react with the thioester and release the free thiol.



Scheme 5.8 E. coli CoA-RNA Capture.

Total RNA from E. coli was captured by thiol sepharose resin (1). Flow-through from column was collected for later (2). Column was washed and CoA-RNA released by addition of DTT (3) as sample A. Flow-through (2) was treated with hydroxylamine to release the free thiol (3) and subjected to another column (5). Flow-through was collected (6) for later. Column was washed and CoA-RNA was released by addition of DTT (7) as thioester-CoA-RNA sample B. Flow-through (6) was subjected to another column, washed, and released (7, 8) as background RNA sample C. Each product A, B, and C was subjected to another column (9, 10, 11), and the final products A', B', and C' denoted by hashed box were processed for NGS library preparation.

The newly generated CoA-RNA, considered to be thioester-CoA-RNA before treatment, was then ethanol precipitated to remove residual salts and DTT before loading to a new thiol sepharose column (Column B), where it was washed and eluted with 100 mM DTT and 100 mM Tris (pH 8) the same as Column A. The initial flow-through of the Column B was retained in order to be

loaded to a third thiol sepharose column (Column C) for similar washes and purification, and would serve as a negative control for subsequent RT-PCR experiments. The RNA purified from thiol Columns A, B, and C were each subjected to a secondary thiol sepharose column to further enrich CoA-RNA from nonspecifically purified RNA (Columns A', B', and C', respectively). The eluted CoA-RNA was ethanol precipitated and recovered in water.

Recovered CoA RNA samples were treated with 16 μM Nudt7 for 1 hour at 37 °C and then ligated to the 5' adapter GL1a. While previous tests with T7Ad had shown promising results, GL1a was of a similar efficiency and a much higher abundance considering it was synthetic and in a 0.2 M stock in contrast to T7Ad being made *in vitro* and in a stock of 30 μM after purification. After RT of the ligated RNA with the Adp16 RT primer using the optimized parameters for SSIV, PCR was carried out using primers PcR1 and Adp2 that are complementary to the 3' adaptor installed by RT and the GL1a 5' adaptor, respectively. The result was a strong band in each sample around 150 bp (Fig 5.10A). This result was peculiar given that only results from Columns A and B should have yielded a PCR product, and the product observed is much higher intensity than expected given previous estimates of CoA- and thioester-CoA-RNA abundance (Kowtoniuk et al. 2009). Given the size of the adapter sequences, the range of RNA isolated by this method was estimated to be 100 bp and below, which did not match previously isolated NAD-RNA sequences of near 300 bp (Cahova et al. 2015).

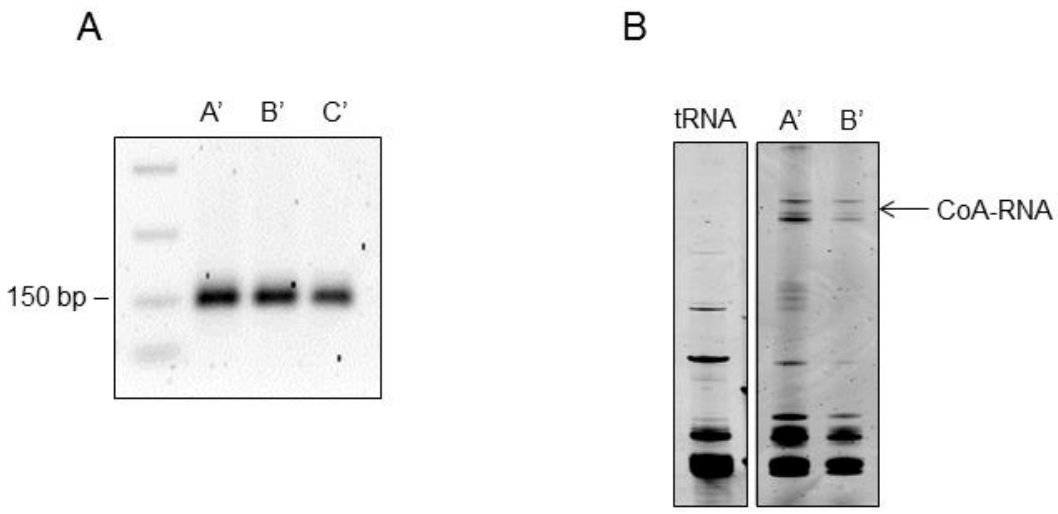


Figure 5.10 RT-PCR on CoA-RNA Capture from E. coli total RNA.

A, RT-PCR of CoA-RNA captured by thiol sepharose column. B, PAGE of captured RNA compared to tRNA.

The isolated RNA was run on 8% denaturing PAGE to visualize the RNA before cDNA and determine if the bands observed from previous PCR were artifacts or if there was in fact isolated CoA-RNA of specific sizes (Fig. 5.10B). RNA from samples A' and B' was easily visualized, and both had similar patterns of migration showing four distinct bands near the front as well as two bands near the wells. There were 4 bands present in the A sample that were not in the B sample that were of low intensity, indicating that the two treatments of the samples did result in isolation of different RNA populations. The intensity of the RNA bands, however, was quite confusing given the picomolar amount of RNA estimated to be isolated. Upon further investigation, similarities of the isolated RNA to tRNA samples were noticed (Fig. 5.10B), and it was concluded that the CoA-RNA capture protocol by the thiol sepharose column had actually captured

thiol-modified tRNA such as 2- and 4-thiouridine (s^2U, s^4U) in addition to CoA-RNA (Fig. 5.11) (Leipuviene et al. 2004). From the gel it appeared that the thiol-modified tRNA was the majority of the RNA captured.

Figure 5.11 Thiol-modified RNA found in tRNA.

Despite the lack of homogeneity produced by the CoA-RNA capture, Figure 5.11B illustrates that RNA of various sizes were isolated that cannot be attributed to tRNA contamination. Previous investigations produced NAD-RNA of around 300 nt in length, which was a consequence of the capture protocol that employed a terminal transferase activity and annealing between few cytosines and guanine for a splint adapter ligation (Kowtoniuk et al. 2009). Such a method has the similar drawbacks to template-switching mechanism, as the annealing hinges upon the terminal cytosines of the transcript and guanosines of the adaptor for success and fidelity. As a result, the relatively short-length sRNA were seen as enriched, and other RNAs that were not previously-annotated sRNA were determined to be 5’ fragments of known mRNA related to

metabolism. Considering that method employed selected for RNA of a certain size, its unknown whether these shortened mRNA fragments occur in this form within the cell or not. Based on Figure 10B, it seems there are larger RNA are slightly larger than the 90 nt tRNA, and some that are significantly larger than tRNA, including a portion that is barely able to enter the 8% polyacrylamide gel. This indicates that CoA could be attached to whole mRNAs and not just 5' terminal fragments. In addition, there is the possibility that CoA-RNA was captured of the same size as tRNA but was masked on the gel due to the intensity of the tRNA itself. Before pursuing the sequencing of these RNAs, it is important to purify them to heterogeneity. Otherwise the sequence reads would be overwhelmingly dominated by the tRNA impurities.

5.3 Conclusions

The CoA-capture protocol was proven effective, but there were still pieces that required optimization. In order to exclude the enrichment of the relatively high concentrations of tRNA in the capture method, the Nudt7 enzyme would be a necessity. Rather than using the thiol sepharose column to discriminate thiol-containing RNA from others, formation of more stable covalent bonds would be required so that Nudt7 hydrolysis could occur and specifically release tRNA. Considering the problems encountered with the capture efficiency of previous iterations of the iodo-functionalized resins tested above, another type of functional group with a more amenable linker would be required for both the CoA-RNA capture and the Nudt7 hydrolysis to be effective.

Due to the previous challenges with the capture of CoA-RNA by the highly reactive iodo functional group affixed to solid state resin, a better system for CoA capture could be designed to be carried out in solution. The use of a biotin-$(PEG)_n$-maleimide molecule would be advantageous for this purpose due to the formation of a thioether bond upon the reaction of the maleimide with the free thiol of CoA in pH ranges lower than allowed for iodo-sulfide conjugation (Fig. 5.12). The lower pH would offer increased stability for the RNA and allowed the increased contact time with the resin to maximize conjugation efficiency. After reaction, the RNA itself could be ethanol precipitated to remove any unreacted linker, and the biotin group can be captured by streptavidin resin. The presence of the linker group between the CoA-RNA and the streptavidin should give enough space for Nudt7 hydrolysis to occur, therefore specifically releasing the CoA-RNA as 5'-pRNA that is readily ligated to a 5' adapter and ready to undergo RT-PCR.

Previous attempts at Nudt7 hydrolysis of resin functionalized with PEG linkers resulted in a low enzyme activity (TGSIA), and it was uncertain if PEG was interacting with the enzyme itself. The use of other organic linkers to create space between substrate and solid surface would create solubility problems, as these RNA captures and enzyme hydrolyses need to occur in aqueous solutions. Therefore PEG, or PEG-like linkers such as TOTDDA (4,7,10-trioxa-1,13-tridecanediamine) used in the ULPCIA resin seem to be a good candidates for this purpose, and the length of the PEG linker will be a crucial component to tune in order for the CoA hydrolysis to be optimized. While the presence of PEG

groups to a certain extent keeps the complex soluble, they greatly affect the solubility of the Biotin-PEG-Maleimide-CoA-RNA complex during ethanol precipitation. The principle of ethanol precipitation is that in high concentrations of salt and in organic solvent such as ethanol, the ionic bonds between the ionic compounds become strong enough that the RNA precipitates from solution. PEG is known to enhance the solubility of organic compounds through the exclusion of water molecules, and is also used in biochemical reactions as a concentrator to increase the interactions of ionic compounds in a large space. The incorporation of PEG linkers at increased unit size can prevent the conjugated RNA from precipitating, therefore preventing the concentration of the captured CoA-RNA. For these reasons, the PEG linker size will need to be optimized for the capture protocols.

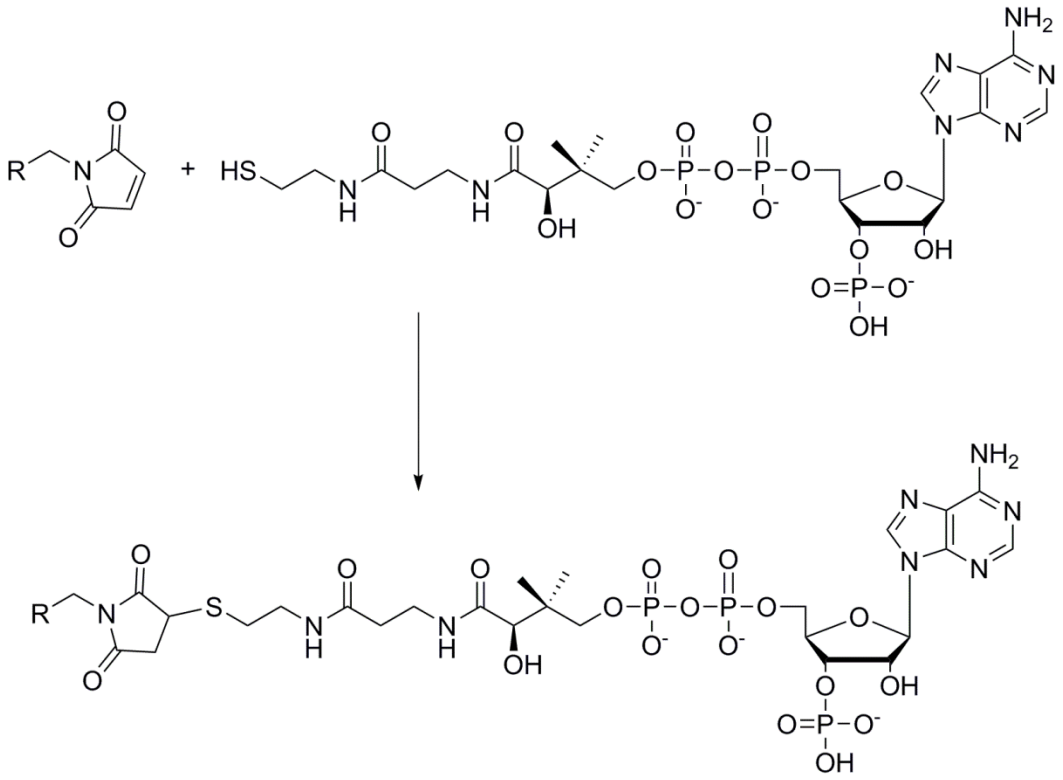


Figure 5.12 CoA-Capture by Maleimide.

The capture of CoA-RNA for identification is a large task with many pitfalls. Based on the number of components involved from beginning to end of this protocol, it seems the lack of CoA-RNA sequence identification reported in the 7 years since its discovery is for good reason. Despite the problems encountered with each step of the way, this work presents multiple pathways for which the CoA-RNA can be captured and identified. As a result, CoA-RNA was successfully captured from *E. coli* total RNA, but purification to homogeneity was not achieved. Although multiple avenues for specifically capturing and releasing CoA-RNA were proven successful, their efficiencies are adequate, at best, and still require some amount of optimization.

CHAPTER VI – PERSPECTIVES

The identification of CoA-RNA is a goal that would uncover many unknown aspects of bacterial metabolism. Despite the significance of the identification of NAD-RNA sequences, so many more questions were realized upon the analysis of the data (Cahova et al. 2015). In addition to the realization that the sequences identified only represented those from one organism grown in one condition, the question of the metabolism of these coenzyme-RNA conjugates is in question.

In terms of CoA-RNA, the degradation can potentially be carried out by NudC or NudL *in vitro*, but how it occurs in the cells is another matter (Bird et al. 2016). Based on kinetic data, both enzymes are not suited for high turnover rates of CoA alone given its cellular concentrations. The K_m of NudL for CoA of 0.5 mM, in particular, is high in comparison to the known concentrations of CoA up to 0.4 mM in the cell (Miller et al. 2007). Even in the event that cells are in nutrient rich conditions and this number is at its peak, the turnover number of NudL is substantially low in comparison to other Nudix CoAses (AbdelRaheim and McLennan 2002). Therefore the existence of NudL as a CoAse within *E. coli* must occur at a very tempered state, where the induction of its expression is required to influence the rate of CoA hydrolysis.

The evolutionary aspect of such a poor enzyme as NudL comes into focus when considering how it could have survived divergence to the γ-Proteobacteria. Given that prokaryotes do not typically possess organelles to compartmentalize cellular processes, the majority of biochemical components of metabolism are

located in the bacterial cytoplasm, if not the cell wall. Considering that CoA instrumental in growth and metabolism, the presence of a high-activity CoA hydrolase in the cell would be detrimental to survival. The eukaryotic counterparts of *C. elegans* NDX8 and *M. musculus* Nudt7 do have a much higher activity, but they also enjoy the benefit of peroxisomal localization via signal peptide targeting. With the main store of CoA existing in the mitochondria, these eukaryotic enzymes are separated by multiple membranes from their substrate, and thus are only capable of interacting with derivatives of CoA that would come within proximity such as oxidized CoA or thioester-CoA. Therefore these eukaryotic enzymes have no selective disadvantage for being highly active within the cell.

NudC is a peculiar enzyme due to its wide-spread substrate ambiguity. When tested *in vitro*, NudC was shown capable of hydrolyzing over 10 different nucleotide-derived substrates with different activities, including 3 common coenzymes in different oxidation states such as NAD (Frick and Bessman 1995). For such an enzyme to exist within *E. coli* in the open world of the cytoplasm with such unbridled targeting of metabolically integral compounds, one would think NudC activity would be inversely proportional to cell survival. The turnover number of NudC is less than the eukaryotic Nudix CoAses, but the ability to cleave important redox molecules such as NAD and FAD, especially with the high affinity for NADH, implies its activity to be dangerous to cell survival. Given the importance of the build-up of reduced NADH to metabolism and ATP production, it’s curious to think that such an enzyme is ready and waiting to

waste all of its potential. The same could be said for CoA, of course, however NADH is not capable of dimer formation via oxidation render it useless, and is also not known for complex metabolism such as fatty acid metabolism that requires an abundance of sequential biochemical reactions for its regeneration. Both of these things are true of CoA and have been speculated as to the purpose of certain Nudix CoAses (Cartwright et al. 2000, Reilly et al. 2008).

If these Nudix hydrolases NudC and NudL exist as a part of CoA-RNA metabolism, this metabolism could not be carried out quickly. Given the amount of CoA-RNA estimated in the cell of 8 fmol per μg RNA, these concentrations would bring both NudC and NudL to extremely low activities. Increasing the enzyme concentrations within the cell through some type of regulation event would increase this activity, but this implies some regulation either on the level of transcription or translation. The role of these Nudix enzymes in bacterial CoA-RNA metabolism needs to be probed by the generation of knockout mutants for each, and the subsequent transformation of a plasmid that encodes each enzyme. This plasmid should be under the control of a dose-dependent inducible promoter such as the arabinose promoter, so that different levels of NudC and NudL activity can be tuned while assessing the quantity of captured CoA-RNA.

The Nudix enzymes cloned were chosen for specific purposes. While the determination of NudL as a specific Nudix CoAse in *E. coli* and the characterization of CoA as a substrate for NudC were goals set out by this study, the foremost reason for cloning these enzymes was to develop a CoA-RNA capture protocol. As of now, the world is without any knowledge of these

sequences. The sequences themselves represent a wealth of knowledge on a topic unknown before 2009. As such, the retrieval and identification of such RNA is vital to the understanding of its purpose.

There is much debate, albeit mostly between the two labs who have published work on the topic and our own, about the generation of CoA-RNA and its purpose within the cell. It is hard to imagine that CoA-RNA exists simply as a cap to RNA as proposed for NAD-RNA. Firstly, the existence of redundant capping structures serves no evolutionary purpose, especially considering that there are other, more available adenosine derivatives that could fulfill this role such as AppA without siphoning indispensable metabolic coenzymes. Second, the installation of these coenzymes into RNA cannot be achieved by the current model of transcription initiation (Bird et al. 2016). Transcription initiation by CoA is out of the question, as is achieving levels of depCoA to out-compete ATP for transcript initiation (Huang 2003a). Third, the existence of the thioester-CoA-RNA implicates the comparably higher complexity of CoA-RNA than NAD-RNA (Kowtoniuk et al. 2009).

Four versions of CoA-RNA were discovered in bacteria, indicating that the CoA can continue to react as an acylation substrate when attached to RNA. The reasoning for this is, again, that CoA cannot be used to initiate transcription, and therefore thioester-CoA cannot either. Therefore the CoA must be installed before thioesterification either as depCoA in some unforeseeable way given the cellular environment, or through a mechanism similar to that of the depCoA-synthesizing ribozyme that utilizes ATP and PhP (Huang et al. 2000).

Considering the data available, the latter mechanism seems more plausible. In addition, the three separate CoA-thioesters found with RNA do not occur in the same organism, meaning the RNA modification could be species specific. Considering this, it is more likely that CoA is installed similar to the ribozyme synthesized mechanism and then acylated based on the specific cellular environment than the possibility that different organisms evolved specific machinery that allow the installation of distinct CoA-thioesters as initiators. Furthermore, if CoA-RNAs can be acylated based on a similar rationale to CoA acylation, it's possible that these modified RNAs are a product of the energy state of the cell. Therefore the identification of CoA-RNA from cultures grown in different conditions that influence the state of the cell and therefore the state of CoA is important to discerning the reasoning behind this modification. In addition, the idea of CoA-RNA in different acylation states is reminiscent of coenzyme-RNA ribozymes hypothesized from the RNA World predating protein enzymes. It is therefore intriguing to consider that sequences may exist in contemporary bacteria that act as ribozymes and facilitate the acylation activity of CoA, and possibly participate in specialized metabolism. Such sequences would never have been discovered before based on limitations in CoA-RNA capturing and sequencing, both of which can now be overcome with modern methods.

The use of any of the studied Nudix hydrolases could surely achieve hydrolysis of CoA in a capture protocol. However, given the problems encountered with both the release of captured CoA-RNA from solid resin and the inherent contamination of non-CoA-RNA, NudC should be used with caution due

to its substrate ambiguity *in vitro*. The specific CoAse Nudt7 is the most viable option at this moment for release due to its ability to hydrolyze CoA-thioesters (Reilly et al. 2008). Capture resins, therefore can be modeled to mimic lauroyl-CoA after capture, as this was the molecule that showed the highest activity with Nudt7. Given the assumptions of the solid resins in suspension with the CoA-RNA inhibiting enzyme degradation, the focus may need to move to a system where hydrolysis can occur in solution after capture and purification on the beads.

The capture of thiol-modified tRNA presents a large problem with the overall capture protocol. These tRNA can participate in every purification step along with CoA-RNA, including the capture by thiol cross-linking chemistries. The cellular abundance of these tRNAs can cause them overtake the true CoA-RNA in amplification for NGS library preparation. Therefore the discriminatory step must be Nudix CoAse release of CoA-RNA from solid resin, as this offers the most accessible method of physically separating molecules that have reacted from those that have not. Ideally, the resin of a high capacity would react with both CoA-RNA and thiol-tRNA, and Nudt7 digestion would allow the CoA-RNA to elute while the thiol-tRNA remains. This requires the development of a resin that is capable of facilitating enzyme hydrolysis at efficiencies greater than the 50% observed in this study.

The enzyme hydrolysis of CoA-RNA could also be increased by a more active CoAse such as *C. elegans* NDX8. However due to the struggles of obtaining a soluble protein from recombinant expression in *E. coli*, other avenues

of generating this enzyme should be investigated. The original investigator who characterized this enzyme did it with a TrxA fusion from a T7 promoter similar to the pET32b system used in this study, however their yields went unreported and were apparently very poor when observing their gels (AbdelRaheim and McLennan 2002). If this enzyme is capable of hydrolyzing CoA-RNA, which is likely considering that Nudt7 had no problem and they are structurally related, it could be used for a more efficient hydrolysis of CoA from resin in a capture protocol. Mammalian expression systems may be an avenue worth pursuing for this endeavor if bacterial expression cannot be achieved.

Considering the entirety of this study, the identities of CoA-RNA remain unknown. The capture of CoA-RNA has been demonstrated, however its lack of purity taints the result that would come from NGS. CoA-RNA metabolism is still largely unknown, however with the data provided here the case for extended regulation of the CoA-RNA pathway can be made. Identification of multiple Nudix enzymes capable of CoA hydrolysis in *E. coli* offer new targets for probing their roles in the CoA-RNA pathway, or the CoA pathway alone considering that CoA degradation has never itself been characterized in *E. coli* (Leonardi et al. 2005). This study presents many viable options for many steps of a protocol to capture and identify of CoA-RNA, and even demonstrates its success. The future directions of this area should continue to be those that strive for stringency and purity of results. The focus should not be towards a specific size range of RNA, but towards all possible sizes of CoA-RNAs to get the true picture of what these RNAs are and how they function within the cell. Only by focusing on the entire

picture and not excluding certain areas can we hope to grasp the understanding of such a process as CoA-RNA function.

REFERENCES

AbdelRaheim, S. R. and A. G. McLennan. The Caenorhabditis elegans Y87G2A.14 Nudix Hydrolase is a Peroxisomal Coenzyme A Diphosphatase. *BMC Biochem.* **2002,** *3*, 5.

Aravin, A. A.; G. J. Hannon and J. Brennecke. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science.* **2007,** *318*, 761-4.

Balibar, C. J.; M. F. Hollis-Symynkywicz and J. Tao. Pantethine rescues phosphopantothenoylcysteine synthetase and phosphopantothenoylcysteine decarboxylase deficiency in Escherichia coli but not in Pseudomonas aeruginosa. *J. Bacteriol.* **2011,** *193*, 3304-12.

Bandyra, K. J. and B. F. Luisi. Licensing and due process in the turnover of bacterial RNA. *RNA Biol.* **2013,** *10*, 627-35.

Bandyra, K. J.; N. Said; V. Pfeiffer; M. W. Gorna; J. Vogel and B. F. Luisi. The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNase E. *Mol. Cell.* **2012,** *47*, 943-53.

Baneyx, F. and M. Mujacic. Recombinant protein folding and misfolding in Escherichia coli. *Nat. Biotechnol.* **2004,** *22*, 1399-408.

Bartel, D. P. and J. W. Szostak. Isolation of New Ribozymes from a Large Pool of Random Sequences. *Science.* **1993,** *261*, 1411-18.

Benner, S. A.; A. D. Ellington and A. Tauer. Modern metabolism as a palimpsest of the RNA world. *Proc. Natl. Acad. Sci. U. S. A.* **1989,** *86*, 7054-58.

Bessman, M. J.; D. N. Fricks and S. F. O'Handley. The MutT Proteins or "Nudix" Hydrolases, a Family of Versatile, Widely Distributed, "Housecleaning" Enzymes. *J. Biol. Chem.* **1996,** *271*, 25059-62.

Bird, J. G.; Y. Zhang; Y. Tian; N. Panova; I. Barvik; L. Greene; M. Liu; B. Buckley; L. Krasny; J. K. Lee; C. D. Kaplan; R. H. Ebright and B. E. Nickels. The mechanism of RNA 5' capping with NAD+, NADH and desphospho-CoA. *Nature.* **2016.**

Bocobza, S. E. and A. Aharoni. Switching the light on plant riboswitches. *Trends Plant Sci.* **2008,** *13*, 526-33.

Boisselier, E. and D. Astruc. Gold nanoparticles in nanomedicine: preparations, imaging, diagnostics, therapies and toxicity. *Chem. Soc. Rev.* **2009,** *38*, 1759-82.

Cahova, H.; M. L. Winz; K. Hofer; G. Nubel and A. Jaschke. NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. *Nature.* **2015,** *519*, 374-7.

Cartwright, J. L.; L. Gasmi; D. G. Spiller and A. G. McLennan. The Saccharomyces cerevisiae PCD1 gene encodes a peroxisomal nudix hydrolase active toward coenzyme A and its derivatives. *J. Biol. Chem.* **2000,** *275*, 32925-30.

Chapman, K. B. and J. W. Szostak. Isolation of a ribozyme with 5'-5' ligase activity. *Chem. Biol.* **1995,** *2*, 325-33.

Chen, X.; N. Li and A. D. Ellington. Ribozyme Catalysis of Metabolism in the RNA World. *Chem. Biodivers.* **2007,** *4*, 633-55.

Chen, Y. G.; W. E. Kowtoniuk; I. Agarwal; Y. Shen and D. R. Liu. LC/MS analysis of cellular RNA reveals NAD-linked RNA. *Nat. Chem. Biol.* **2009,** *5*, 879-81.

Chiang, P. K.; R. K. Gordon; J. Tal; G. C. Zeng; B. P. Doctor; K. Pardhasaradhi and P. P. McCann. S-Adenosylmethionine and methylation. *FASEB J.* **1996,** *10*, 471-80.

Coleman, T. and F. Huang. RNA-Catalyzed Thioester Synthesis. *Chem. Biol.* **2002,** *9*, 1227-36.

Coleman, T. and F. Huang. Optimal Random Libraries for the Isolation of Catalytic RNA. *RNA Biol.* **2005,** *2*, 129-36.

Coleman, T.; G. Wang and F. Huang. Superior 5' homogeneity of RNA from ATP-initiated transcription under the T7 phi-2.5 promoter. *Nucleic Acids Res.* **2004,** *32*, e14.

Coleman, T. M.; N. Li and F. Huang. A simple and efficient method to prepare thioesters in aqueous solutions. *Tetrahedron Letters.* **2005,** *46*, 4307-10.

Dong, H.; J. Lei; L. Ding; Y. Wen; H. Ju and X. Zhang. MicroRNA: function, detection, and bioanalysis. *Chem. Rev.* **2013,** *113*, 6207-33.

Dvinge, H.; E. Kim; O. Abdel-Wahab and R. K. Bradley. RNA splicing factors as oncoproteins and tumour suppressors. *Nat. Rev. Cancer.* **2016.**

Ekland, E. H. and D. P. Bartel. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature.* **1996,** *382*, 373-76.

Fontecave, M.; M. Atta and E. Mulliez. S-adenosylmethionine: nothing goes to waste. *Trends Biochem. Sci.* **2004,** *29*, 243-9.

Frick, D. N. and M. J. Bessman. Cloning, Purification, and Properties of a Novel NADH Pyrophosphatase. *J. Biol. Chem.* **1995,** *270*, 1529-34.

Garten, A.; S. Petzold; A. Korner; S. Imai and W. Kiess. Nampt: linking NAD biology, metabolism and cancer. *Trends Endocrinol. Metab.* **2009,** *20*, 130-8.

Gasmi, L. and G. McLennan. The Mouse Nudt7 Gene Encodes a Peroxisomal Nudix Hydrolase Specific for Coenzyme A and its Derivatives. *J. Biochem.* **2001,** *357*, 33-38.

Geerlof, A.; A. Lewendon and W. V. Shaw. Purification and Characterization of Phosphopantethiene Adenylyltransferase from Escherichia coli *J. Biol. Chem.* **1999,** *274*, 27105-11.

Giancaspero, T. A.; G. Busco; C. Panebianco; C. Carmone; A. Miccolis; G. M. Liuzzi; M. Colella and M. Barile. FAD synthesis and degradation in the nucleus create a local flavin cofactor pool. *J. Biol. Chem.* **2013,** *288*, 29069-80.

Gilbert, W. The RNA World. *Nature.* **1986,** *319*, 618.

Grushoff, P. S.; S. Shany and A. W. Bernheimer. Purification and Properties of Streptococcal Nicotinamide Adenine Dinucleotide Glycohydrolase. *J. Bacteriol.* **1975,** *122*, 599-605.

Guan, C. D.; P. Li; P. D. Riggs and H. Inouye. Vectors the facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. *Gene.* **1988,** *67*, 21-30.

Ha, M. and V. N. Kim. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell. Biol.* **2014,** *15*, 509-24.

Hornung, V.; J. Ellegast; S. Kim; K. Brzozka; A. Jung; H. Kato; H. Poeck; S. Akira; K.-K. Conzelmann; M. Schlee; S. Endres and G. Hartmann. 5'-Triphosphate is the Ligand for RIG-I. *Science.* **2006,** *341*, 994-97.

Huang, C.; M. Li; C. Chen and Q. Yao. Small interfering RNA therapy in cancer: mechanism, potential targets, and clinical applications. *Expert Opin. Ther. Targets.* **2008,** *12*, 637-45.

Huang, F. Efficient incorporation of CoA, NAD and FAD into RNA by in vitro transcription. *Nucleic Acids Res.* **2003a,** *31*, e8.

Huang, F. Efficient incorporation of CoA, NAD, and FAD into RNA by in vitro transcription. *Nucleic Acids Res.* **2003b,** *31*, e8.

Huang, F.; C. W. Bugg and M. Yarus. RNA-Catalyzed CoA, NAD, and FAD Synthesis from Phosphopantetheine, NMN, and FMN. *Biochemistry.* **2000,** *39*, 15548-55.

Huang, F. and Y. Shi. Synthesis of symmetrical thiol-adenosine conjugate and 5' thiol-RNA preparation by efficient one-step transcription. *Bioorg. Med. Chem. Lett.* **2010,** *20*, 6254-7.

Huang, F. and Y. Shi. Synthesis of photolabile transcription initiators and preparation of photocleavable functional RNA by transcription. *Bioorg. Med. Chem. Lett.* **2012,** *22*, 4254-8.

Huang, F.; G. Wang; T. Coleman and N. Li. Synthesis of adenosine derivatives as transcription initiators and preparation of 5' fluorescein- and biotin-labeled RNA through one-step in vitro transcription. *RNA.* **2003,** *9*, 1562-70.

Huang, F. and M. Yarus. 5'-RNA Self-Capping from Guanosine Diphosphate. *Biochemistry.* **1997,** *36*, 6557-63.

Illangasekare, M.; G. Sanchez; T. Nickles and M. Yarus. Aminoacyl-RNA Synthesis Catalyzed by an RNA. *Science.* **1995,** *267*, 643-47.

Jackowski, S. and C. O. Rock. Regulation of Coenzyme A Biosynthesis. *J. Bacteriol.* **1981,** *148*, 926-32.

Jackowski, S. and C. O. Rock. Metabolism of 4'-Phosphopantetheine in Escherichia coli. *J. Bacteriol.* **1984,** *158*, 115-20.

Jadhav, V. R. and M. Yarus. Acyl-CoAs from Coenzyme Ribozymes. *Biochemistry.* **2002,** *41*, 723-29.

Jenny, R. J.; K. G. Mann and R. L. Lundblad. A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. *Protein Expression and Purification.* **2003,** *31*, 1-11.

Jinek, M.; K. Chylinksi; I. Fonfara; M. Hauer; J. A. Doudna and E. Charpentier. A Programmable Dual-RNA-Guided DNA Endouclease in Adaptive Bacterial Immunity. *Science.* **2012,** *337*, 816-21.

Joosten, V. and W. J. van Berkel. Flavoenzymes. *Curr. Opin. Chem. Biol.* **2007,** *11*, 195-202.

Joyce, G. F. RNA evolution and the origins of life. *Nature.* **1989,** *338*, 217-24.

Kanehisa, M.; M. Araki; S. Goto; M. Hattori; M. Hirakawa; M. Itoh; T. Katayama; S. Kawashima; S. Okuda; T. Tokimatsu and Y. Yamanishi. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **2008,** *36*, D480-4.

Kang, L. W.; S. B. Gabelli; M. A. Bianchet; W. L. Xu; M. J. Bessman and L. M. Amzel. Structure of a Coenzyme A Pyrophosphatase from Deinococcus radiodurans: a Member of the Nudix Family. *J. Bacteriol.* **2003,** *185*, 4110-18.

Kim, J.; J. P. Kershner; Y. Novikov; R. K. Shoemaker and S. D. Copley. Three serendipitous pathways in E. coli can bypass a block in pyridoxal-5'-phosphate synthesis. *Mol. Syst. Biol.* **2010,** *6*, 436.

Ko, J. H. and Y. Lee. RNA-conjugated template-switching RT-PCR method for generating an Escherichia coli cDNA library for small RNAs. *J. Microbiol. Methods.* **2006,** *64*, 297-304.

Kowtoniuk, W. E.; Y. Shen; J. M. Heemstra; I. Agarwal and D. R. Liu. A chemical screen for biological small molecule-RNA conjugates reveals CoA-linked RNA. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106*, 7768-73.

Kumar, R. K. and M. Yarus. RNA-Catalyzed Amino Acid Activation. *Biochemistry.* **2001,** *40*, 6998-7004.

Kupke, T.; J. A. Caparrós-Martín; K. J. Malquichagua Salazar and F. A. Culiáñez-Macià. Biochemical and physiological characterization of Arabidopsis thaliana AtCoAse: a Nudix CoA hydrolyzing protein that improves plant development. *Physiologia Plantarum.* **2009,** *135*, 365-78.

LaVallie, E. R.; E. A. DiBlasio; S. Kovacic; K. L. Grant; P. F. Schendel and J. M. McCoy. A Thioredoxin Gene Fusion Expression System That Circumvents Inclusion Body Formation in the E. coli Cytoplasm. *Bio/Technology.* **1993,** *11*, 187-93.

Leipuviene, R.; Q. Qian and G. R. Bjork. Formation of Thiolated Nucleosides Present in tRNA from Salmonella enterica serovar Typhimurium Occurs in Two Principally Distinct Pathways. *J. Bacteriol.* **2004,** *186*, 758-66.

Leonardi, R. and S. Jackowski. Biosynthesis of Pantothenic Acid and Coenzyme A. *EcoSal. Plus.* **2007,** *2.*

Leonardi, R.; Y. M. Zhang; C. O. Rock and S. Jackowski. Coenzyme A: back in action. *Prog. Lipid. Res.* **2005,** *44*, 125-53.

Leonardo, M. R.; Y. Dailly and D. P. Clark. Role of NAD in Regulating the adhE Gene of Escherichia Coli. *J. Bacteriol.* **1996,** *178*, 6013-18.

Li, N. and F. Huang. Ribozyme-Catalyzed Aminoacylation from CoA Thioesters. *Biochemistry.* **2005,** *44*, 4582-90.

Li, N.; C. Yu and F. Huang. Novel cyanine-AMP conjugates for efficient 5' RNA fluorescent labeling by one-step transcription and replacement of [gamma-32P]ATP in RNA structural investigation. *Nucleic Acids Res.* **2005,** *33*, e37.

Lin, H. Nicotinamide adenine dinucleotide: beyond a redox coenzyme. *Org. Biomol. Chem.* **2007,** *5*, 2541.

Lin, S.-J. and L. Guarente. Nicotinamide adenine dinucleotide, a metabolic regulator of transcription, longevity and disease. *Curr. Opin. Cell Biol.* **2003,** *15*, 241-46.

Lohse, P. A. and J. W. Szostak. Ribozyme-catalysed amino-acid transfer reactions. *Nature.* **1996,** *381*, 442-44.

Lorsch, J. R. and J. W. Szostak. In vitro evolution of new ribozymes with polynucleotide kinase activity. *Nature.* **1994,** *371*, 31-36.

Louie, T. M.; X. S. Xie and L. Xun. Coordinated Production and Utilization of FADH2 by NAD(P)H-Flavin Oxidoreductase and 4-Hydroxyphenylacetate 3-Monooxygenase. *Biochemistry.* **2003,** *42*, 7509-17.

Malakhov, M. P.; M. R. Mattern; O. A. Malakhova; M. Drinker; S. D. Weeks and T. R. Butt. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genomics.* **2004,** *5*, 75-85.

Malygin, A. G. and M. F. Shemyakin. Adenosine, NAD and FAD can Initiate Template-Dependent RNA Synthesis Catalyzed by Escherichia coli RNA Polymerase. *FEBS Letters.* **1979,** *102*, 52-55.

Marchler-Bauer, A.; M. K. Derbyshire; N. R. Gonzales; S. Lu; F. Chitsaz; L. Y. Geer; R. C. Geer; J. He; M. Gwadz; D. I. Hurwitz; C. J. Lanczycki; F. Lu; G. H. Marchler; J. S. Song; N. Thanki; Z. Wang; R. A. Yamashita; D. Zhang; C. Zheng and S. H. Bryant. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **2015,** *43*, D222-6.

Marraffini, L. A. and E. J. Sontheimer. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.* **2010,** *11*, 181-90.

Matz, M.; D. Shagin; E. Bogdanova; O. Britanova; S. Lukyanov; L. Diatchenko and A. Chenchik. Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Res.* **1999,** *27*, 1558-60.

McLennan, A. G. The Nudix hydrolase superfamily. *Cell Mol. Life. Sci.* **2006,** *63*, 123-43.

McLennan, A. G. Substrate ambiguity among the nudix hydrolases: biologically significant, evolutionary remnant, or both? *Cell Mol. Life. Sci.* **2013,** *70*, 373-85.

Mildvan, A. S.; Z. Xia; H. F. Azurmendi; V. Saraswat; P. M. Legler; M. A. Massiah; S. B. Gabelli; M. A. Bianchet; L. W. Kang and L. M. Amzel. Structures and mechanisms of Nudix hydrolases. *Arch. Biochem. Biophys.* **2005,** *433*, 129-43.

Miller, J. R.; J. Ohren; R. W. Sarver; W. T. Mueller; P. de Dreu; H. Case and V. Thanabal. Phosphopantetheine adenylyltransferase from Escherichia coli: investigation of the kinetic mechanism and role in regulation of coenzyme A biosynthesis. *J. Bacteriol.* **2007,** *189*, 8196-205.

Miller, S. L. and G. Schlesinger. Prebiotic Syntheses of Vitamin Coenzymes: II. Pantoic Acid, Pantothenic Acid, and the Composition of Coenzyme A. *J. Mol. Evol.* **1993,** *36*, 308-14.

Milligan, J. F.; D. R. Groebe; G. W. Witherell and O. C. Uhlenbeck. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **1987,** *15*, 8783-98.

Munafo, D. B. and G. B. Robb. Optimization of enzymatic reaction conditions for generating representative pools of cDNA from small RNA. *RNA.* **2010,** *16*, 2537-52.

Nakamura, T.; T. Pluskal; Y. Nakaseko and M. Yanagida. Impaired coenzyme A synthesis in fission yeast causes defective mitosis, quiescence-exit failure, histone hypoacetylation and fragile DNA. *Open Biol.* **2012,** *2*, 120117.

O'Handley, S. F.; D. N. Frick; C. A. Dunn and M. J. Bessman. Orf186 Represents a New Member of the Nudix Hydrolases, Active on Adenosine(5')triphospho(5')adenosine, ADP-ribose, and NADH. *J. Biol. Chem.* **1998,** *273*, 3192-97.

Pan, T. and O. C. Uhlenbeck. A small metalloribozyme with a two-step mechanism. *Nature.* **1992,** *358*, 560-63.

Papenfort, K. and J. Vogel. Regulatory RNA in bacterial pathogens. *Cell Host Microbe.* **2010,** *8*, 116-27.

Parrish, S.; W. Resch and B. Moss. Vaccinia virus D10 protein has mRNA decapping activity, providing a mechanism for control of host and viral gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104*, 2139-44.

Reilly, S. J.; V. Tillander; R. Ofman; S. E. Alexson and M. C. Hunt. The Nudix Hydrolase 7 is an Acyl-CoA Diphosphatase Involved in Regulating Peroxisomal Coenzyme A Homeostasis. *J. Biochem.* **2008,** *144*, 655-63.

Romaniuk, E.; L. W. McLaughlin; T. Neilson and P. J. Romaniuk. The Effect of Acceptor Oligoribonucleotide Sequence on the T4 RNA Ligase Reaction. *Eur. J. Biochem.* **1982,** *125*, 639-43.

Sengle, G.; A. Eisenfuhr; P. S. Arora; J. S. Nowick and M. Famulok. Novel RNA catalysts for the Michael reaction. *Chem. Biol.* **2001,** *8*, 459-73.

Song, M. G.; S. Bail and M. Kiledjian. Multiple Nudix family proteins possess mRNA decapping activity. *RNA.* **2013,** *19*, 390-9.

Song, M. G.; Y. Li and M. Kiledjian. Multiple mRNA decapping enzymes in mammalian cells. *Mol. Cell.* **2010,** *40*, 423-32.

Szutowicz, A.; H. Bielarczyk; A. Jankowska-Kulawy; T. Pawelczyk and A. Ronowska. Acetyl-CoA the key factor for survival or death of cholinergic neurons in course of neurodegenerative diseases. *Neurochem. Res.* **2013,** *38*, 1523-42.

Tarasow, T. M.; S. L. Tarasow and B. E. Eaton. RNA-catalysed carbon-carbon bond formation. *Nature.* **1997,** *389*, 54-57.

Tsukiji, S.; S. B. Pattnaik and H. Suga. An alcohol dehydrogenase ribozyme. *Nat. Struct. Biol.* **2003,** *10*, 713-7.

Tucker, B. J. and R. R. Breaker. Riboswitches as versatile gene control elements. *Curr. Opin. Struct. Biol.* **2005,** *15*, 342-8.

Tuerk, C. and L. Gold. Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science.* **1990,** *249*, 505-210.

Tuschl, T.; P. A. Sharp and D. P. Bartel. Selection in vitro of novel ribozymes from a partially randomized U2 and U6 snRNA library. *EMBO J.* **1998,** *17*, 2637-50.

Unrau, P. J. and D. P. Bartel. RNA-catalysed nucleotide synthesis. *Nature.* **1998,** *395*, 260-63.

Vallari, D. S. and S. Jackowski. Biosynthesis and Degradation Both Contribute to the Regulation of Coenzyme A Content in Escherichia coli. *J. Bacteriol.* **1988,** *170*, 3961-66.

Wagner, E. G. H. and R. W. Simons. Antisense RNA Control in Bacteria, Phages, and Plasmids. *Ann. Rev. Microbiol.* **1994,** *48*, 713-42.

Walsh, C. Flavin Coenzymes: At the Crossroads of Biological Redox Chemistry. *Acc. Chem. Res.* **1980,** *13*, 148-55.

White, H. B. Coenzymes as Fossils of an Earlier Metabolic State. *J. Mol. Evol.* **1976,** *7*, 101-04.

Wilson, C. and J. W. Szostak. In vitro evolution of a self-alkylating ribozyme. *Nature.* **1995,** *374*, 777-82.

Wilson, R. C. and J. A. Doudna. Molecular mechanisms of RNA interference. *Annu. Rev. Biophys.* **2013,** *42*, 217-39.

Wu, D.; P. Hugenholtz; K. Mavromatis; R. Pukall; E. Dalin; N. N. Ivanova; V. Kunin; L. Goodwin; M. Wu; B. J. Tindall; S. D. Hooper; A. Pati; A. Lykidis; S. Spring; I. J. Anderson; P. D'Haeseleer; A. Zemla; M. Singer; A. Lapidus; M. Nolan; A. Copeland; C. Han; F. Chen; J. F. Cheng; S. Lucas; C. Kerfeld; E. Lang; S. Gronow; P. Chain; D. Bruce; E. M. Rubin; N. C. Kyrpides; H. P. Klenk and J. A. Eisen. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature.* **2009,** *462*, 1056-60.

Xu, W.; C. A. Dunn and M. J. Bessman. Cloning and characterization of the NADH pyrophosphatases from Caenorhabditis elegans and Saccharomyces cerevisiae, members of a Nudix hydrolase subfamily. *Biochem. Biophys. Res. Commun.* **2000,** *273*, 753-8.

Yarus, M. Boundaries for an RNA World. *Curr. Opin. Chem. Biol.* **1999,** *3*, 260-67.

York, A. W.; Y. Zhang; A. C. Holley; Y. Guo; F. Huang and C. L. McCormick. Facile synthesis of multivalent folate-block copolymer conjugates via aqueous RAFT polymerization: targeted delivery of siRNA and subsequent gene suppression. *Biomacromolecules.* **2009,** *10*, 936-43.

Zhuang, F.; R. T. Fuchs and G. B. Robb. Small RNA expression profiling by high-throughput sequencing: implications of enzymatic manipulation. *J. Nucleic Acids.* **2012,** *2012*, 360358.