Investigating molecular interactions with 3'-5' RNA polymerases in *Dictyostelium discoideum*

Research Thesis

Presented in partial fulfillment of the requirements for graduation with Research Distinction in Biochemistry in the undergraduate colleges of The Ohio State University

by

Madison Allegretti

The Ohio State University
May 2024

Project Advisor: Dr. Jane E. Jackman, Department of Chemistry and Biochemistry

Abstract

The Thg1/TLP family of enzymes catalyze a number of non-canonical 3'-5' nucleotide addition reactions across all domains of life. They play important roles in tRNA processing reactions such as $tRNA^{His}$ maturation and 5' end repair of truncated tRNAs. Thg1, first discovered in yeast, has a strong preference to add a G_{-1} nucleotide to the 5' end of $tRNA^{His}$ in a non-Watson--Crick base pair, permitting subsequent aminoacylation with histidine. TLPs, or Thg1-like-proteins, kinetically prefer to repair truncated RNA substrates in a Watson-Crick template dependent manner and demonstrate activity on many tRNA species beyond $tRNA^{His}$. This work focuses on two TLP enzymes found in the eukaryotic model organism *Dictyostelium discoideum*: DdiTLP3 and DdiTLP4. DdiTLP3 is targeted to the mitochondria of the cell where it repairs the 5' end of mitochondrial tRNA editing substrates as part of an incompletely understood mechanism. The true in vivo function of cytosolic DdiTLP4 remains unclear but it is an essential enzyme that has demonstrated polymerase activity on a variety of RNA substrates in vitro. Because of the complexity of tRNA editing and the diversity of the reactions catalyzed by Thg1/TLP enzymes encoded in *D. discoideum,* we hypothesize that DdiTLP3 and DdiTLP4 may interact with one or more cellular macromolecules in order to fulfill their respective biological functions. By identifying interacting molecules that facilitate their reactions, unknown mechanistic details of their 3'-5' nucleotide addition reactions can be elucidated. Encoding a FLAG epitope tag to a terminus of these DdiTLPs allows the capture of complexes of FLAG-tagged DdiTLP3 and DdiTLP4 with interacting RNA and proteins that participate in their reactions, which can then be identified with sequencing or mass spectrometry techniques. After concluding that the presence of a FLAG-tag does not impede on the activity of each TLP compared to wild-types, a protocol was optimized for *in vitro* crosslinking and immunoprecipitation (CLIP) with FLAG-tagged DdiTLP4 and total RNA isolated from *D. discoideum*. The efficiencies and specificities of two crosslinking methods, chemical and UV, were compared. *In vivo* CLIP with FLAG-tagged DdiTLP3 and DdiTLP4 in *D. discoideum* will further validate *in vitro* RNA target

identification and also allow purification of interacting proteins. Together, these *in vitro* and *in vivo* approaches will enable identification of interacting RNAs and proteins that are important for the function of DdiTLP3 and DdiTLP4, thus providing insight into the mechanism and function of these unusual 3'-5' polymerases. By further understanding the interactions that occur between DdiTLP3 and DdiTLP4 and other candidate molecules, we can understand the exact roles these enzymes play in *D. discoideum* and possibly apply our findings to the roles of TLPs in other eukaryotes where their function has not yet been demonstrated.

Introduction

The field of molecular biology is centered upon a central dogma universal to all domains of life positing that genetic information stored in DNA is decoded to synthesize proteins, which facilitate biological reactions vital to keeping cells alive. With the DNA biopolymer consisting of nucleic acid monomers and protein consisting of amino acid monomers, there is a need for an intermediary that can "translate" the chemical information from the "language" of DNA into that of proteins. RNA fulfills this role in the central dogma, completing the directional transformation of information in DNA by transcription into messenger RNA (mRNA), which is then translated into protein, using transfer RNAs (tRNAs) that decode the mRNA message and add amino acids to a growing polypeptide chain.

Each species of tRNA is unique in their biochemical characteristics which are designed to carry a certain amino acid and recognize a specific type of codon. Fidelity and specificity of the translation process is vital to the survival of the cell and is tightly regulated by a multitude of highly conserved, intertwined processes that ensure that the genetic instructions from DNA are properly carried through to the final polypeptide products. Successful translation relies on the ability of a given tRNA species to both be nonspecifically recognized by the ribosome and very specifically recognized by its cognate aminoacyl-tRNA synthetase that "charges" it with its respective amino acid. These contrasting needs are met by evolutionary pressure to maintain the L-shaped tertiary structure of tRNA which allows universal recognition by the ribosome, alongside chemical and structural identity elements specific to each tRNA species that promote selective recognition by their corresponding synthetase. For example, primary sequence motifs can serve as identity elements for accurate aminoacylation, as well as unique post-transcriptional chemical modifications to the tRNA species. (Figure 1) (Sprinzl et al. 1998; Marck and Grosjean 2002). Some of these post-transcriptional modifications are nearly ubiquitous across all tRNA species in each domain of life, such as 5' and 3' leader sequence removal, while others are specific to certain types of tRNA species (Gerber and Keller, 1999; Jackman, 2010). A universally conserved identity

element unique to $tRNA^{His}$ is the presence of a G_{-1} at its 5' end that is crucial for recognition by histidyl-tRNA synthetase (Sprinzl et al 1998; Himeno et al. 1989; Nameki et al 1995; Giegé et al 1998; Rosen and Musier-Forsyth 2003). In eukaryotes, this G_{-1} is added post-transcriptionally by $tRNA^{His}$ guanylyltransferase, or Thg1, demonstrating a polymerase activity that seemingly contradicts the well-established 5'-3' directionality of nucleotide polymerization (Figure 2) (Cooley et al. 1982; Gu et al. 2003). Interestingly, this G_{-1} is added in a non-templated fashion across from an A_{73} discriminator nucleotide, forming a non-Watson-Crick base pair and raising more questions about the catalytic ability of the Thg1 polymerase (Nameki et al. 1995). Further investigations into the catalysis of 3'-5' polymerization found that it employs the same two-metal-ion mechanism as 5'-3' polymerization, with the Thg1 polymerase active site showing a striking structural similarity to the active site of canonical polymerases (Figure 3) (Jackman and Phizicky, 2006; Smith and Jackman, 2012; Hyde et al 2010).

Phylogenetic analysis of the Thg1 protein sequence revealed Thg1-related genes encoded in bacterial and archaeal species that genetically encode the G_{-1} nucleotide on $tRNA^{His}$ that is retained after 5' end processing, leaving no apparent necessity for a G_{-1} addition enzyme (Figure 4) (Burkard and Söll 1988; Burkard et al 1988; Rao et al. 2010). Biochemical characterization of some of these Thg1 family members revealed their ability to catalyze broader 3'-5' nucleotide addition reactions on tRNAs beyond just $tRNA^{His}$, with catalytic abilities distinct from those of Thg1: *in vitro* they prefer more to repair truncated tRNA substrates than to add N_{-1} residues to full-length tRNAs, and they add nucleotides in a templated Watson-Crick base pair manner (Rao et al. 2011). These Thg1-related proteins thus were classified as “Thg1-like-proteins” or TLPs for their phylogenetic and biochemical distinctions (Figure 4). The preference of TLPs to restore the 5' end of tRNA opens up the possibility of their role in tRNA 5' repair activities, but much remains to be discovered regarding this potential function (Rao et al. 2010). Due to the lack of a Thg1 gene in any bacteria or archaea, it is theorized that TLPs represent an ancestral

3'-5' polymerase family that gave rise to the $tRNA^{His}$-specific Thg1 in eukaryotes, although some lower eukaryotes have retained one or more TLPs (Jackman et al. 2012).

Dictyostelium discoideum is a eukaryotic slime mold that happens to encode four Thg1 family members, known as DdiTLP1/DdiThg1, DdiTLP2, DdiTLP3, and DdiTLP4, all which have distinct cellular localizations and biological functions, and will be the focus of this work (Figure 5) (Abad et al. 2011; Long et al. 2016). By expressing both a Thg1 and three TLPs, this organism represents an evolutionary intermediate between the Thg1-lacking bacteria and archaea and the TLP-lacking multicellular eukaryotes, providing opportunity for deeper study of the evolutionary history and biological capabilities of TLPs. Furthermore, the sequenced *D. discoideum* genome is readily available and the organism proves to be biochemically tractable; thus, we choose it as a model organism to study the Thg1/TLP enzyme family.

The biological functions of three of the Thg1/TLP enzymes in *D. discoideum* have been demonstrated. Sequence alignment and nucleotide addition activity assays revealed that DdiTLP1 is localized to the cytosol and adds a G_{-1} to the 5' end of $tRNA^{His}$, homologous to the activity of the *Saccharomyces cerevisiae* Thg1 (yThg1) that spearheaded the discovery of the Thg1/TLP family of enzymes, thus earning the homologous name DdiThg1 (Abad et al. 2011). DdiTLP2 is targeted to the mitochondria, where it similarly adds a G_{-1} to the 5' end of $tRNA^{His}$ but specifically recognizes its mitochondrial substrate using a different mechanism than the GUG-codon recognition utilized by DdiThg1 and yThg1 in the cytosol (Long et al. 2016; Johnecheck et al. unpublished). DdiTLP3 is also targeted to the mitochondria, where it participates in a vital cellular process known as 5' mitochondrial 5' tRNA editing. Mitochondrial 5' tRNA editing, discovered in the protist *Acanthamoeba castellanii*, is a tRNA modification pathway common amongst some eukaryotic microbes (Figure 6) (Lonergan and Gray 1993). 5' editing is necessary in these species because some of their mitochondrial tRNA (mt-tRNA) species are encoded with mismatched nucleotides within the first three positions in their amino acyl-

acceptor stem, which are excised by a nuclease before a 3'-5' polymerase activity restores the correct base-pairs (Lonergan and Gray 1993, Price and Gray 1999). DdiTLP3 is the first documented candidate of the 5' tRNA editing enzyme, demonstrating 5' repair activity on known mt-tRNA repair substrates in *D. discoideum* (Figure 6) (Abad et al 2011, Long et al 2016). However, the complete mechanism and molecular players of mitochondrial 5' tRNA editing are not known. The identity of the nuclease that acts in the first step of the reaction remains unknown, as well as the nature of its potential complex with DdiTLP3. Additionally, the 5' repair enzyme in any other species, likely a homolog of DdiTLP3, remains to be identified, outlining potential roles of uncharacterized TLPs in these other eukaryotes. Viable DdiTLP3 deletion strains of *D. discoideum* cannot be obtained and RNA-interference (RNAi) depletion strains show significant developmental defects, demonstrating the need for mitochondrial tRNA 5' editing by DdiTLP3 in this organism (Long et al. 2016).

The true biological role of DdiTLP4 remains a mystery, but it catalyzes 5' repair activity i*n vitro* on a variety of RNA substrates, including those of DdiTLP3, introducing the possibility of DdiTLP4 participating in similar 5' editing functions (Long et al. 2016). However, DdiTLP4 is localized to the cytoplasm of the cell, where there is no current knowledge of a role for 5' end repair for which DdiTLP4 could play a part in. Previous attempts at characterizing the biological role of DdiTLP4 presented numerous obstacles and yielded somewhat inconclusive results. Despite its elusive role in the cell, DdiTLP4 is known to be essential to the viability of the organism, shown by the inability to isolate deletion strains for its associated gene and the significant growth defects when its gene expression is depleted using RNAi (Long et al. 2016). 5S ribosomal RNA (rRNA), snoRNAs, and the Class I non-coding RNA (ncRNA) species unique to *D. discoideum* are previously identified candidate RNA substrates of DdiTLP4; however, the abundance of these RNAs in the cell cytoplasm likely allowed their ease of detection in the biochemical assays we used to identify them—we observe multiple other less-abundant potential RNA substrates in our assays that remain unidentified, thus preventing our conclusion of the

general kinds of RNAs DdiTLP4 prefers to act on (Long et al. 2016; S. Dodbele dissertation, 2019). Class I ncRNAs are known to play key roles in the development of *D. discoideum* throughout its life cycle, thus a model is proposed where loss of DdiTLP4 processing of these vital ncRNAs leads to the phenotype of DdiTLP4-depleted strains (S. Dodbele dissertation, 2019). However, attempts to pinpoint the exact 5' end state of an RNA target of DdiTLP4 have been unsuccessful. Sequenced 5' ends of RNAs found to have differential abundances in wild-type versus DdiTLP4-depleted *D. discoideum* strains show varying distributions of nucleotides at the 5' end, making *in vitro* recapitulation of *in vivo* activity very difficult due to inability to clone one discrete substrate (S. Dodbele dissertation, 2019). Furthermore, it is important to consider that highly modified RNAs may evade many sequencing-based methods due to the inability of reverse transcriptase (RT) used for RNA-seq to read over modifications, which could cause highly modified ncRNAs that DdiTLP4 may act on to be excluded from sequencing libraries, further complicating biochemical characterization. The heterogeneity of 5' ends of potential DdiTLP4 substrates suggests that this enzyme may have a broader substrate specificity for processing various types of ncRNAs, including snoRNAs and Class I developmental RNAs.

From its transcription in the nucleus to its docking in the ribosome, a tRNA species is involved in numerous pathways through which it must interact with a unique, complex set of cellular machinery to ensure its stability and functionality during translation (Houser-Scott and Engelke, 2001). Figure 1B shows some examples of pathways a tRNA species may be involved in during its lifetime and some of the necessary interactions that must occur with each. Many of these molecular players, including those that participate in the tRNA modification reactions of the TLP enzymes, remain unknown, such as the nuclease in the mt-tRNA 5' editing reaction in *Dictyostelium discoideum* (Figure 6). Moreover, the detection of non-tRNA substrates of DdiTLP4 brings into consideration the potential role that TLPs play in modifying other types of ncRNAs and the biological implications of those pathways, thus the types of RNAs that DdiTLP4 interacts with is especially of interest. Due to the complexity of tRNA post-

transcriptional modification reactions and the diversity of the reactions catalyzed by the different TLPs in *D. discoideum*, we hypothesize that each enzyme interacts with a unique set of cellular macromolecules in order to carry out their biological functions. In this work we aim to identify proteins and RNAs that participate in the RNA modification reactions catalyzed by DdiTLP3 and DdiTLP4, which will provide further insight into the unknown mechanistic details of their respective functions in the cell. To accomplish this, we employ an affinity purification method targeting an epitope tag fused to our DdiTLP, allowing us to pull down our target protein complexed with associating molecules that can be subsequently identified. Here we show that DdiTLP3 and DdiTLP4 are amenable to epitope tagging with a FLAG tag and describe our optimization of protocols suitable to perform *in vitro* crosslinking and immunoprecipitation (CLIP) with our enzymes, laying the groundwork for a novel method to characterize their biological activity.

D. discoideum is the only species in which discrete biological roles for TLPs have been defined, so further biochemical characterization of these enzymes in this model organism will lay the groundwork for understanding the potential roles of the Thg1/TLP family members in other organisms, as well as provide more insight into the evolution of the 3'-5' polymerases from TLPs to Thg1. Due to evidence that these TLPs are vital to survival of *D. discoideum*, their characterization can help us understand the biological importance of 3'-5' polymerases across other eukaryotes, including in other organisms where mt-tRNA editing takes place and even in humans where only the Thg1 polymerase remains encoded in the genome.

Chapter 1: Engineering epitope-tagged DdiTLP3 and DdiTLP4 for use in crosslinking and immunoprecipitation (CLIP)

In order for a tRNA species to efficiently carry out its role in translation, it must undergo a unique series of post-transcriptional modifications that renders its structure able to be non-specifically recognized by a translating ribosome and specifically recognized by its cognate aminoacyl-tRNA synthetase. The accurate translation of a protein is vital for it to be able to fold properly and subsequently carry out its biological function; thus, the cell invests a significant amount of energy in maintaining homeostasis of the translation process, including the transcription and processing of tRNA species, which are the most abundant nucleic acids in a cell on a molar basis (Waldron and Lacroute, 1975; Jackman, 2010). In fact, between 1-10% of a given genome of any species codes for enzymes involved in tRNA modification pathways, and a number of these enzymes are ancient and highly conserved across the domains of life (Yacoubi et al. 2012; Jackman, 2010). Furthermore, loss of certain tRNA modifications in humans have been linked with disease, and we have observed that Thg1/TLP deletion strains of our model organisms are not viable (Abott et al 2014, Long et al 2016). These findings underscore the importance of tRNA modification enzymes across all life. Thus, there is much interest in studying the cellular machineries involved in vital tRNA modification pathways. Due to the interconnectedness of these pathways and close physical proximity of some of their steps, it is reasonable to believe that proteins and RNAs involved in certain reactions may be interacting as a complex, a common phenomenon observed in subcellular-compartmentalized metabolic pathways (Schmitt and An, 2017). Identifying all the molecular players of a tRNA modification pathway can provide useful insight into its complete mechanism(s), including details about substrate recognition and specificity, catalysis, any participating cofactors, and potential downstream targets, to name a few.

We aim to identify protein and RNA molecules that interact with DdiTLP3 and DdiTLP4 during their enzymatic reactions to further understand the mechanistic details of their modification pathways.

In the mt-tRNA 5' editing reaction catalyzed by DdiTLP3, a nuclease acts in the first step to cleave mismatched nucleotides at the 5' end of certain tRNA species. Following the aforementioned logic of metabolic compartmentalization, it is reasonable to believe that DdiTLP3 forms a complex with this nuclease to perform its complementary biological function. In the case of DdiTLP4, whose cellular function remains a mystery, identifying its interacting proteins and RNAs can provide us insight to what tRNA (or other ncRNA) processing pathway it may play a role in, including what kind of RNA substrates it prefers to act on.

We aim to utilize the method of crosslinking and immunoprecipitation (CLIP) to co-purify complexes of our TLP enzymes with interacting molecules. The immunoprecipitation approach to protein purification takes advantage of the highly specific binding that occurs between an antibody and its specific antigen by immobilizing an antibody specific to the target protein on a suitable bead resin, acting as “bait” that will “capture” the protein target. Since the discovery of the method, variations and optimizations of it have been established for use in specific contexts across biological disciplines to affinity purify proteins and investigate co-purifying molecules. For example, rather than using a protein-specific antibody to capture protein targets, many researchers choose to utilize recombinant epitope-tagged constructs of their protein of study, which can be targeted with antibodies against the tag sequence rather than the protein itself, allowing a more universal type of affinity tag (Fritze et al. 2000). This use of a discrete target binding sequence at the N- or C- terminus of a protein allows highly specific immunoprecipitations with low nonspecific background (Gerace and Moazed, 2015). The FLAG peptide sequence (DYKDDDDK) is an especially useful epitope tag due to its relatively small size that usually avoids interfering with the structure of its fused protein, as well as the commercial availability of anti-FLAG antibodies (Brizzard, 2008; Hernan et al. 2018); thus, we chose the FLAG epitope tag to fuse to our DdiTLP enzymes of study. Here we describe the cloning, expression, and purification of FLAG-tagged

DdiTLP3 and DdiTLP4 constructs, followed by biochemical activity assays to determine if the presence of a FLAG tag impacts their 5' addition activity.

Results

Cloning, expression, and purification of FLAG-tagged DdiTLP3 and DdiTLP4 constructs

We aimed to create constructs of DdiTLP3 and DdiTLP4 fused to a FLAG epitope tag for use in CLIP to identify molecules that interact with these proteins during their enzymatic reactions. Canonical use of the FLAG tag involves the genetic insertion of its nucleic acid sequence to the to the N- or C- termini of a protein gene of interest and subsequent expression and purification from a suitable vector transformed into *E. coli* competent cells (Figure 7). In this work, I cloned, expressed and purified three FLAG-TLP constructs. I designed oligonucleotide primers to insert the FLAG sequence at the desired terminus of each protein gene, and polymerase chain reaction (PCR) with Phusion site-directed mutagenesis amplified AVA421 vectors containing each DdiTLP gene fused to the FLAG insert. Successful addition of the FLAG sequence was confirmed with Sanger sequencing (data not shown). I overexpressed and purified the FLAG-TLP proteins from Rosetta pLysS *E. coli* competent cells. Wild-type DdiTLP3 and DdiTLP4 were simultaneously expressed and purified to be used downstream as controls. Resulting protein concentrations were measured with the BioRad Bradford protein assay (Bradford, 1976), ranging from 0.6-3.1 mg/mL. The purity of proteins was assessed with SDS-PAGE, all estimated to be >90% pure except for DdiTLP3 which is around 50% (Figure 8a).

Immunoblotting of FLAG-tagged DdiTLP3 and DdiTLP4

To confirm the presence of the FLAG epitope tag on purified mutant DdiTLP3 and DdiTLP4 protein constructs, I carried out western blot using anti-FLAG primary antibodies. Western blot analysis confirmed an abundant presence of a FLAG tag on each of our DdiTLP constructs of interest due to the

presence of a strong signal in FLAG-tag positive lanes at the expected molecular weight of each protein, using a molecular weight ladder as reference. The signal is dependent on the amount of protein loaded onto the gel, further confirming that we are detecting FLAG-tagged TLPs (Figure 8b).

Determination of the catalytic efficiency of FLAG-tagged DdiTLP3 and DdiTLP4

We sought to confirm whether the presence of a FLAG tag on a terminus of our DdiTLP enzymes of interest impacts their ability to carry out their catalytic activity compared to wild-type untagged protein. We investigate this with a phosphatase protection activity assay (PPA), first developed by Jackman and Phizicky, suitable for analyzing the 5' tRNA repair propensity of TLP enzymes (Figure 9a) (Jackman and Phizicky, 2006). I performed phosphatase protection assays with my purified FLAG-DdiTLP enzymes and a ^{32}P-labeled mt-tRNA 5' editing substrate from *D. discoideum* missing one nucleotide at its 5' end. I first qualitatively assessed activity by comparing amounts of end-point products made in reactions with FLAG-TLP versus wild-type controls, and later determined single-turnover kinetic parameters for each DdiTLP. Qualitative assessment of the relative signal intensity of end-point products of FLAG-tagged DdiTLPs versus wild-types led us to conclude that the FLAG-tagged DdiTLP3 and DdiTLP4 perform their 5' nucleotide addition reactions with comparable efficiency to wild-types, although C-terminal FLAG DdiTLP4 appears to reach a larger maximal product formation than N-terminal FLAG DdiTLP4 (Figure 9b). For this reason, I selected C-terminal FLAG-tagged DdiTLP4 over the N-terminal FLAG for primary use in subsequent assays for experimental simplicity, although either construct would have sufficed. Performing triplicate (n=3) time-point assays for each DdiTLP (n=2 for C-terminal FLAG DdiTLP4) and fitting data to a single-exponential rate equation allowed me to calculate their apparent reaction rates k_{obs} (Figure 9c, d). The DdiTLP3 wild-type and FLAG mutant both have a k_{obs} around 0.3 min^{-1}, although of note is the somewhat large error range associated with C-terminal FLAG DdiTLP3. As for DdiTLP4, the

calculated rate of the C-terminal FLAG construct, 0.85 min^{-1}, appears to be two-fold higher than that of the wild-type, 0.41 min^{-1}, but of note is the large error range associated with C-terminal FLAG DdiTLP4.

Discussion

In this chapter, we describe the cloning, expression, and purification of epitope-tagged DdiTLP3 and DdiTLP4. We selected the FLAG epitope tag to fuse to our DdiTLP proteins of study for highly specific purification of complexes of each DdiTLP with molecules that interact during their enzymatic reactions. Three FLAG-TLP constructs were created: C-terminal FLAG DdiTLP3, C-terminal FLAG DdiTLP4, and N-terminal FLAG DdiTLP4. We decided to create both N- and C-terminal FLAG-tagged versions of DdiTLP4 to compare the catalytic activities of each and decide which terminal the tag is most stable on. As for DdiTLP3, only a C-terminal FLAG construct was created because a mitochondrial-targeting peptide tag (mito-tag) is already encoded onto the N-terminus in its wild-type gene, and previous attempts at cloning a FLAG-tag next to a mito-tag on the N-terminus of DdiTLP2 led to defects in enzyme activity (A. Nguyen, senior thesis, 2020). I designed oligonucleotide primers to insert the FLAG sequence at the desired terminus of each DdiTLP gene. Each DdiTLP gene was previously cloned into an AVA421 vector containing an N-terminal His_6 tag for affinity chromatography purification, an ampicillin resistance gene (beta-lactamase) and an IPTG-inducible T7 promoter (Abad et al. 2011). Due to the somewhat lengthy 24 base-pair FLAG nucleic acid sequence, forward and reverse primers were created with one half of the insertion sequence on each of their 5' ends. Phusion PCR was chosen as our method of site-directed mutagenesis due to its ability to introduce point mutations, deletions, and insertions to double-strand DNA (dsDNA) plasmids. Correct use of this method involves phosphorylating each primer at their 5' ends so that when the mutated DNA plasmid is PCR amplified, the linear DNA can be easily ligated back to a circular plasmid conformation. I treated each PCR reaction with DpnI restriction enzyme to digest remaining parent plasmids. Analyzing the product of each DpnI digestion versus negative controls on

agarose gels allowed me to qualitatively determine if the PCR reaction occurred, because supercoiled parent plasmid DNA will run farther on the gel than the amplified linear DNA (data not shown). Following treatment of each successful reaction with DNA ligase, the plasmids were transformed using heat-shock into competent BL21 *E. coli* cells to be further copied and subsequently purified with a miniprep. Transformation is another quality checkpoint of the cloning process, because the proliferation of transformed colonies suggests that the cloned plasmid was properly synthesized and thus able to express beta-lactamase to ensure growth on LB plates supplemented with ampicillin. Each purified TLP plasmid was sent to Sanger sequencing to confirm successful addition of the FLAG sequence. I experienced no notable difficulties in cloning the FLAG tag onto each DdiTLP construct, which preliminarily suggested that DdiTLP3 and DdiTLP4 are amenable to affinity-tagging on the termini we chose.

Next, I transformed each purified plasmid containing the FLAG-TLP gene into Rosetta pLysS *E. coli* competent cells to be subsequently purified using immobilized metal-ion affinity chromatography targeting the His_6 tag also encoded on the recombinant protein. I simultaneously expressed and purified wild-type DdiTLP3 and DdiTLP4 to be used downstream as controls. Resulting protein concentrations ranged from 0.6-3.1 mg/mL. Of note is the relatively low concentration of wild-type DdiTLP3 at 0.6 mg/mL. Furthermore, SDS-PAGE analysis of pre-and post- induction samples of DdiTLP3 cell culture showed relatively low expression of DdiTLP3 even before purification was attempted (data not shown). The purity of proteins was assessed with SDS-PAGE, estimated by judging the relative intensity of Coomassie-stained target protein bands at their approximate molecular weight in kilodaltons (kDa) versus other present bands. All proteins were estimated to be >90% pure except for DdiTLP3, which is around 50% (Figure 8a). Since DdiTLP3 co-purified with other protein contaminants, it is likely that the measured concentration overestimates the actual concentration of DdiTLP3 in the stock. The low yield and purity of wild-type DdiTLP3 could raise concern for the utility of this protein as a control in

downstream assays; however, previous work with DdiTLP3 in the lab encountered similar challenges with achieving high expression, yield, and purity, but still observed its robust activity in *in vitro* biochemical assays, thus we chose to move forward with these results. It is interesting to note the ~5-fold higher yield of C-terminal FLAG-tagged DdiTLP3 over the wild-type and its higher purity. There is no obvious explanation for why FLAG-tagged DdiTLP3 would have greater expression than the wild-type; however, one possible caveat is that the FLAG-tagged DdiTLP plasmids were cloned and transformed into *E. coli* competent cells fresh for purification, while the wild-type plasmid was purified from a storage strain kept for many years at -80°C before transformation into an *E. coli* expression strain; perhaps there are storage-related defects to the wild-type plasmid. Our successful purifications of FLAG-tagged DdiTLP3 and DdiTLP4 suggest that epitope-tagged versions of DdiTLP3 and DdiTLP4 are readily expressed in *E. coli* competent cells and present no obstacles to metal-ion affinity purification utilizing their N-terminal His_6 tag.

Next, I sought to confirm the presence of the FLAG epitope tag on purified mutant DdiTLP3 and DdiTLP4 protein constructs with western blot. This analysis method was chosen due to its high specificity and sensitivity in detecting a target peptide. Although the FLAG sequence was present on each plasmid as confirmed with Sanger sequencing, it remains a possibility that while each FLAG-DdiTLP construct is expressed and purified, it adopts an aberrant folded conformation that could lead to degradation of the protein and loss of the FLAG tag. However, western blot results signified that a large amount of FLAG tag exists in each FLAG-TLP sample, allowing us to conclude that each purified FLAG-TLP construct retains the FLAG tag at its terminus.

Next, I determined whether the presence of the FLAG tag on the terminus of each DdiTLP would impair its catalytic activity. Although the FLAG peptide is rather small and thus thought not to hinder the proper function of its fused protein, we have seen defects in DdiTLP enzyme activity potentially caused by presence of the FLAG tag (A. Nguyen, senior thesis, 2020). If the FLAG tag interferes with catalytic

function, it is highly likely that it is interfering with the proper folding and/or potential catalytic residues in the protein, and thus could impact the binding of other proteins and RNAs that may participate in its reaction. In the context of our CLIP approach, this impediment could lead to nonspecific immunoprecipitation or its failure altogether. The phosphatase protection assay was employed to test the activity of each FLAG-tagged DdiTLP with wild-types as controls. The premise of this assay involves labeling the 5' end of a truncated tRNA species with a radioactive ^{32}P phosphate group and incubating it with a TLP, which if catalytically active can act to replace the missing nucleotides. Following RNase digestion of the product tRNA into identifiable fragments, the reactions are treated with a phosphatase, and resolved with thin-layer chromatography (TLC). If the TLP was able to perform the 5' repair reaction, the radioactive phosphate group is “protected” by the added nucleotides and thus cannot be cleaved off by the phosphatase, leading to larger-sized tRNA fragments that run farther on the TLC plate (Figure 9). *D. discoideum* mt-tRNA^{Trp} with a G_{+1} 5' start, a known mt-tRNA 5' repair substrate in *D. discoideum*, was chosen as the substrate for this assay in the interest of simplified product analysis since only 1 nucleotide needs to be added to repair. If the FLAG-tagged TLP can properly repair the substrate, following RNase digestion we should observe up to two product spots that run higher on the TLC plate. The highest spot corresponds to the end product of the nucleotidyl-transfer reaction (A_{+1}*pG_{+2}), while the next-highest product spot corresponds to the adenylated intermediate of the Thg1/TLP 5' addition mechanism (Ap*pG_{+2}) (Jahn and Pande, 1991). RNase T1, which cleaves after the 3' phosphate of guanosine residues, was used to quench the reactions. I first carried out activity assays with serial dilutions of each of our enzymes to the endpoints of their reactions, running for two hours before quenching. End-point assays run in this fashion allowed us to qualitatively compare the 5' nucleotide-addition efficiency of each FLAG-tagged TLP to their wild-type counterparts. In these analyses, I concluded that the efficiency of each FLAG-tagged TLP is similar to their wild-type, judging from the relative summed intensities of the signals of the two possible products. While both N-terminal and C-

terminal FLAG DdiTLP4 produce about at least as much product as the wild-type DdiTLP4, it appears as though C-terminal FLAG DdiTLP4 produces a higher yield of total product than N-terminal FLAG. For this reason, I decided to move forward with the C-terminal FLAG over the N-terminal FLAG for simplicity in subsequent kinetic assays and CLIP studies, although either FLAG-tagged protein would have still fared well. Interestingly, it appears as though C-terminal FLAG DdiTLP4 has a higher turnover than even the wild-type, suggesting that the FLAG tag could be enhancing its activity in some way.

Conducting time-point activity assays under single-turnover conditions and determining kinetic parameters allows us to quantitatively compare the efficiencies of each FLAG-tagged TLP to their wild-type counterparts. Triplicate trials were performed for each enzyme, except for C-terminal FLAG DdiTLP4 for which only two trials were performed. This inconsistency is due to the time restraint with performing all trials with all enzymes before the radioactive label on the substrate decays; a third trial was in fact performed with C-terminal FLAG DdiTLP4, but the substrate's radioactivity was too decayed to produce viable data. Time-point data was fitted to a single-exponential rate equation to calculate the observed reaction rates k_{obs} for each DdiTLP. The apparent calculated rate of the wild-type DdiTLP3 is about 1.5-fold higher than the C-terminal FLAG, but of note is the somewhat large error range associated with C-terminal FLAG DdiTLP3. Considering the uncertainty associated with this measurement of k_{obs}, as well as that associated with wild-type DdiTLP3, there are regions of overlap between the error ranges where it is possible that the k_{obs} are more similar or the same. However, we must also note the possible lower bound of the error of the C-terminal FLAG DdiTLP3 k_{obs}, where it could be possible that the efficiency of this enzyme is slightly lower than that of the wild-type. Nevertheless, with this data we can still conclude that the activity of FLAG-tagged TLP3 isn't significantly impaired compared to the wild-type. As for the DdiTLP4 enzymes, we see that the apparent calculated rate of C-terminal FLAG DdiTLP4 is over two-fold higher to that its wild-type counterpart, however of note is the large error range associated with C-terminal FLAG DdiTLP4. The observation that C-terminal FLAG

DdiTLP4 has a higher efficiency than the wild-type DdiTLP4 is consistent with our qualitative analysis of the end-point assays, but the uncertainty with the measurement makes us unable to confidently make that conclusion. However, even considering the uncertainties associated with each measurement, the k_{obs} of C-terminal FLAG DdiTLP4 is at its least similar or the same as the wild-type, if not higher. The potential higher efficiency of FLAG-tagged DdiTLP4 is only advantageous to us in our downstream CLIP studies as it could mean the enzyme binds its substrate more tightly or with more ease, allowing us to easily identify which RNAs it prefers to act upon. Overall, more trials should be performed with both sets of enzymes to potentially minimize the uncertainties of each measurement and thus to make a more informed conclusion of their comparative activities. Furthermore, I can attempt to better fit my time-point data to different kinetic models to potentially calculate more precise rates. However, despite the limitations that were described above with the activity assay data obtained in this work, we can conclude that the presence of a FLAG tag on a terminus of DdiTLP3 and DdiTLP4 does not significantly impact its 5' nucleotide addition activity. These studies allow us to perform our planned CLIP studies without concerns that the FLAG epitope tag impacts the folded TLP conformation and thus the ability of interacting molecules to bind while they perform their biological functions.

Chapter 2: Optimizing protocols and analysis methods for in vitro CLIP with FLAG-tagged DdiTLP3 and DdiTLP4

In Chapter 1 we determined that epitope-tagged versions of DdiTLP3 and DdiTLP4 can be expressed and purified from *E. coli* competent cells, and that the presence of a FLAG tag on one of their termini does not significantly affect their ability to catalyze 5' nucleotide addition reactions. With this knowledge in hand, we can utilize our FLAG-tag TLPs to perform immunoprecipitations (IPs) to identify molecules that interact with our enzymes during their reactions.

The use of a FLAG epitope tag in IPs is advantageous for our purposes over other types of IP methods for several reasons. First, its 8-amino acid sequence DYKDDDDK is highly immunogenic, meaning that high-affinity antibodies can be raised against it. Further, agarose bead resins containing the anti-DYKDDDDK sequence are commercially available that can specifically target the FLAG sequence, allowing specific detection and easy purification of proteins tagged with the epitope. The anti-FLAG sequence on the beads thus acts as the "bait" to capture a FLAG-tagged construct, and when purifying from cell extracts containing other molecules, the FLAG-tagged protein can be co-purified with molecules that bind to and interact with it (Figure 10). Various analysis methods have been established to identify IP co-purifying molecules, including microarray hybridization, mass spectrometry, and next-generation nucleic acid sequencing techniques. We hope to identify specifically bound molecules that co-purify with DdiTLP3 and DdiTLP4 in order to further understand their biological functions and mechanisms in *D. discoideum.*

During IP, stringent salt-washes are utilized to eliminate non-specific protein and RNA interactions with the protein of interest and enrich for specifically interacting complexes. However, if the protein-protein or protein-RNA complex is not stabilized by a covalent bond, which is the case for the many transient interactions that occur with RNA-binding proteins (RBPs) like our TLPs, stringent wash conditions risk washing away specific binding partners that have lower affinity. Furthermore,

during IP pull-down experiments without crosslinking, it is possible that RNA/DNA and protein binding partners can reassociate with different, non-specific binding partners and then when pulled down, produce false positive results (Wheeler et al, 2017). Thus, crosslinking before immunoprecipitation (CLIP) was developed to combat this issue, allowing more stringent washing of a covalently linked protein-RNA or protein-protein complex and ensuring that specific binding partners are pulled down.

There are several commonly used reagents to covalently crosslink an interacting molecule with a target protein, two of which will be utilized in this work: UV light and formaldehyde. Each reagent possesses advantages and disadvantages depending on experimental purpose and context. UV-light produces non-reversible covalent crosslinks via a free-radical mechanism and depends on the inherent photo-reactivity of nucleotide bases, especially pyrimidines, with any amino acid residue of a protein, although certain residues are significantly more susceptible to photo-activation than others (like Cys, Lys, Phe, Trp, and Tyr) (Wheeler et al. 2017). Thus, UV-light can only crosslink proteins to RNAs in very close proximity ("zero-distance" crosslinker), and cannot crosslink protein to protein (Greenberg, 1979). This property allows for highly specific protein-RNA crosslinks that allow for IP purification of RBP-RNA complexes; however, UV crosslinking is a low-efficiency method, often producing crosslink yields as low as <1-5% (Wheeler et al. 2017). This low efficiency may require very sensitive analytical techniques downstream but does ensure that successful crosslinks are with specific binding partners over non-specific ones. On the other hand, formaldehyde, which preferentially crosslinks nucleophilic lysine residues with carbonyl addition chemistry, can reversibly crosslink both proteins to nucleic acids and proteins to proteins (Ramanathan et al. 2019). It works with a significantly higher efficiency than UV, with an ability to crosslink molecules interacting at slightly further distances than the zero-distance UV crosslinker. This is because formaldehyde bridges the crosslinks it mediates, versus UV light directly photoactivating two residues to fuse them together. This property of formaldehyde leads to higher occurrence of crosslinking with non-specific binding partners, like accessory proteins that may bind to

the direct nucleic acid-binding protein(s) of interest, which can complicate identification of direct protein-nucleic acid interactions (Ule et al. 2005).

The traditional IP method has evolved extensively since its development, being adapted to suit various contexts of protein-nucleic acid interactions and experimental conditions both *in vitro* and *in vivo*, making available well-established protocols for methods like CLIP, chromatin-IP (ChIP) RNA-immunoprecipitation (RIP), and photoactivatable ribonucleoside CLIP (PAR-CLIP), to name a few. Advanced CLIP methods often are coupled with next-generation sequencing of nucleic acids and mass spectrometry of proteins to identify pulled-down interacting molecules. However, no *in vitro* or *in vivo* CLIP protocol with our TLP enzymes had been previously established in the lab. In this work we sought to first establish optimal *in vitro* CLIP protocols and analysis methods using our FLAG-tagged DdiTLP3 and DdiTLP4, comparing the specificities and efficiencies of UV light and formaldehyde crosslinking reagents to decide which method is most suitable for our purposes. We first seek to establish *in vitro* CLIP protocols before moving *in vivo* for several reasons. First, our *in vitro* method would pose greater experimental simplicity and higher throughput compared to *in vivo* methods that require careful consideration and addressing some experimental challenges for application *D. discoideum* cell culture. Further, materials for efficient in vitro CLIP were readily available to us, including our purified recombinant FLAG-tagged DdiTLPs and stored stocks of total RNA isolated from *D. discoideum* grown under different conditions. Performing *in vitro* CLIP would allow us to obtain useful preliminary data from our FLAG-DdiTLP CLIP system while simultaneously screening for potential biological RNA-binding partners, a function that would prove most useful for our investigations into DdiTLP4, the only DdiTLP for which true biological RNA substrates remain to be discovered. Thus, in this work we primarily used C-terminal FLAG DdiTLP4 to develop *in vitro* CLIP protocols and analysis methods. Here, we describe protocols optimized to perform *in vitro* CLIP with FLAG-tagged DdiTLP3 and DdiTLP4, and our plans for identification of interacting RNAs with next-generation sequencing methods. We derive our protocols

mainly from available RIP and CLIP protocols in the literature, mainly those by Mukherjee et al (2021), Fuentes-Iglesias et al (2020), Nostrand et al (2016) and the Anti-DYKDDDDK magnetic agarose product manual by ThermoFisher (2017). Most available CLIP protocols are performed *in vivo*, so it was necessary to adapt their methods to our *in vitro* approach. By establishing optimal *in vitro* CLIP methodology for our FLAG-tagged DdiTLPs, we can identify molecules that interact with these 3'-5' polymerases during their RNA processing reactions and further understand their RNA repair pathways.

Results

Bead binding assays establish efficacy of IP with FLAG-tagged DdiTLP3 and DdiTLP4

Before developing protocols for CLIP with our FLAG-tagged TLPs, it was important to ensure that our FLAG/anti-FLAG IP system performed efficiently with our chosen anti-FLAG bead resin, producing highly specific IPs from specific binding to the FLAG epitope tag on each TLP with low non-specific background. An important component of this was to establish buffer conditions optimal for FLAG-TLP binding to anti-FLAG beads, as well as wash buffer conditions that would stringently remove non-specific binding partners while maintaining the specific immunogenic interaction between FLAG and anti-FLAG antibodies. Buffer optimization assays were performed by Brandon Iwaniec, leading us to our choice binding buffer that contains 25mM HEPES, 10mM $MgCl_2$, and 100mM NaCl, and a wash buffer consisting of 1000mM NaCl in PBS. I performed *in vitro* bead-binding assays without RNA or crosslinking reagent present to examine the efficacy of IP for each enzyme. The bead-binding assay procedure was adapted from Mukherjee et al (2021), Fuentes-Iglesias et al (2020), and the Anti-DYKDDDDK magnetic agarose product manual by ThermoFisher (2017). Fractions of each step of the IP were collected, and samples were analyzed with SDS-PAGE versus wild-type controls, and an amount of each purified protein stock equal to the input amount of protein for IP was loaded onto the gels for the purpose of estimating IP

yield (Figure 11). In the elution lanes, I observed more intense Coomassie-stained protein bands at the expected size of each protein when performing the IP with FLAG-tag DdiTLPs than with untagged wild-type protein, observing yields of ~30-40% by estimating by eye the relative band intensity of input protein versus IP elution at the DdiTLP expected molecular weight. Furthermore, in the FLAG-tagged TLP elution lanes, the band corresponding to the FLAG-DdiTLP is darker than other bands present (note: the bands present at ~50 kDa and ~25 kDa are the heavy and light chains from denaturing the anti-FLAG antibodies during elution; these are ubiquitous in IP elutions), suggesting that the anti-FLAG agarose is specifically binding with the FLAG epitope tag and that our wash protocol removes potential impurities. Conversely, for wild-type DdiTLP4, I observe a bolder band corresponding to the protein in the flow-through lane than in the elution lane. In other words, FLAG-tagged DdiTLP4 immunoprecipitated with higher yield than wild-type using anti-FLAG magnetic agarose. Of note is the very weak band in the DdiTLP4 purified protein stock lane; some protein was lost during sample preparation; however, the signal of the strong band at 37 kDa in the flow-through lane combined with that of very light band in the elution lane approximate the initial input protein amount, and we can still draw our conclusions based on the relative amounts in each lane. I would expect to see this same result for wild-type DdiTLP3; however, I do not observe a band corresponding to it in the flow-through nor the elution lane. This could be for several reasons. I treated flow-through samples with trichloroacetic acid (TCA) to precipitate proteins out of solution and concentrate them to be easily observed on SDS-PAGE. It is possible that the TCA precipitation was defective for the DdiTLP3 flow-through sample. Moreover, DdiTLP3 purified with a low yield as described previously, and a high volume of stock solution would have been needed to load 40ug of it into a reaction. To prevent too high of a perturbance to our buffer conditions, since our protein dialysis buffer contains glycerol and DTT, we only input about 13ug of wild-type DdiTLP3. Thus, the lack of visible bands in the flow-through could be because so little protein was input initially and thus became heavily diluted in the high-volume wash, unable to be recovered by TCA

precipitation. However, despite some limitations identified in the analysis of these data, the success of elution with FLAG-tagged DdiTLP3 still assures us that the FLAG-tag is efficiently IP'd with this system.

Optimizing *in vitro* CLIP with formaldehyde as a crosslinking reagent

Here we describe the *in vitro* CLIP protocols and analysis methods I developed using formaldehyde as a crosslinking reagent. The *in vitro* protocols are adapted from mostly *in vivo* protocols in the literature, especially those by Hoffman et al (2015), Valasek et al (2007), and the ChIP protocol by Sigma-Aldrich, as well as the previously stated sources.

As previously stated, I mostly use use C-terminal FLAG DdiTLP4 (DdiTLP4 CTF) in my optimization assays. I incubate the FLAG-tagged protein with total RNA isolated from *D. discoideum.* I sought to optimize the amount of input protein, input RNA, bead volume, concentration of formaldehyde used to crosslink, and methods for observing crosslinked RNA. To evaluate the extent of crosslinking in reactions with our enzymes, I crosslinked DdiTLP4-RNA complexes with 1% formaldehyde and observed the results with SDS-PAGE and western blot to evaluate changes in migration of the crosslinked protein that might indicate successful crosslinking of complexes, although whether these complexes contain RNA or protein is not indicated by this assay (Figure 12A). We observe additional diffuse protein bands at higher molecular weights in samples treated with crosslinking reagent compared to the untreated purified protein sample. We do see non-specific background signals, especially at higher molecular weights, in the DdiTLP4 CTF purified protein stock lane, which may correspond to contaminating co-purifying proteins, showing that steps to block non-specific antibody binding during the immunoblotting procedure were insufficient. We then tested two methods to reverse formaldehyde crosslinks to qualitatively analyze immunoprecipitated RNA bound to DdiTLP4 CTF, comparing the methods of proteinase K digestion and heat reversal followed by PCA extraction of RNA. To observe IP'd RNA, I employed radioactive labeling of the 5' end of any extracted RNAs with ^{32}P, visualizing with urea-PAGE. I

more consistently was able to observe RNA after treatment with proteinase K, so decided upon this method for use in future experiments. I proceeded to perform optimization assays with varying concentrations of formaldehyde from 0-1% to investigate the appropriate concentration for our CLIP experiments and observe that 1% formaldehyde produces the most obvious and extensive crosslinking, using western blot and urea-PAGE to visualize immunoprecipitated RNA (Figure 12B). Throughout these optimization processes, I experimented with varying concentrations of input beads, protein, and RNA to maximize the yield of extracted RNA for best visualization with urea-PAGE (data not shown). I determined that inputting as little as 10ug of RNA, 30-50ug of FLAG-tagged enzyme, and the appropriate amount of beads for maximum binding capacity results (60-80uL) produced IP RNA samples that can be visualized with urea-PAGE. However, these amounts can be scaled up to obtain higher yields; bold bands were more consistently observed in assays with higher inputs of around 40-50ug RNA and 100-150ug protein.

***In vitro* CLIP with UV as a crosslinking reagent**

Protocols and analysis methods for *in vitro* UV CLIP with FLAG-tagged TLPs were initially developed by Brandon Iwaniec. The protocol was adapted from available *in vivo* UV CLIP protocols available in the literature, especially those by Mukherjee et al (2021), Nostrand et al (2016), the *in vitro* method described by Götse et al (2021), and the generous advice from Dr. Beth Grayhack (University of Rochester) and Dr. Guramrit Singh (The Ohio State University). Here we outline our results from performing and optimizing *in vitro* UV CLIP with C-terminal FLAG-tagged DdiTLP4.

Rather than reversing the TLP-crosslink and purifying the immunoprecipitated RNA on its own for analysis such as shown in Figure 12B, I applied a UV-CLIP method that first radioactively label the 5' end of RNAs crosslinked to TLPs after IP while the complexes are still associated with the anti-FLAG beads.

This enabled me to resolve the TLP-RNA complex by running the products on an SDS-PAGE gel, and subsequently transferring to a nitrocellulose membrane. The radioactive signal on the membrane detected by a phosphorimager would in principle only correspond to labeled RNA that persisted throughout this analysis because it was crosslinked to a TLP. Before labeling with γ-^{32}P ATP and T4 polynucleotide kinase, I tested the addition of RNase A to the immunoprecipitated samples, which has the potential to generate more uniform RNA fragments crosslinked to the protein, thus enabling better resolution on a gel. When treating samples with a gradient of RNase A concentrations, we observe an RNase A-dependent band that becomes more discrete with increased amounts of RNase A (Figure 13). We scale up the input RNA and protein when performing UV CLIP compared to formaldehyde CLIP due to the low efficiency of UV as a crosslinker, allowing us to still visualize crosslinked complexes. We find that inputting 30-50ug of RNA with 100-150ug of protein produces enough crosslinked complexes to be visualized with radioactive labeling.

Discussion

In this chapter we describe in vitro CLIP protocols optimized for use with FLAG-tagged TLPs. We first performed bead-binding assays with FLAG-tagged DdiTLP3 and DdiTLP4 to test the buffer conditions and IP protocol that we adapted from literature sources, as well as to test the efficacy of IP with FLAG-tagged DdiTLP3 and DdiTLP4 and anti-FLAG magnetic agarose beads. In our optimization stages of CLIP, we use Pierce anti-DYKDDDDK magnetic agarose by ThermoFisher, which is advertised to produce highly specific IPs with greater purity and yields than some other commercially available anti-FLAG beads. Moreover, its magnetic character allows more rapid purification by easy separation with a magnetic stand, rather than lengthier centrifugation steps involved with non-magnetic agarose. Our binding buffer containing HEPES, $MgCl_2$, and NaCl contains many components of the previously established buffer optimized for TLP activity assays which contains HEPES, $MgCl_2$, NaCl, DTT, and BSA. In the context

of our CLIP methodology, however, DTT has the ability to denature the immobilized antibodies on anti-FLAG agarose, while BSA is an exogenous protein that could possibly interfere with specific binding events with our TLPs. Buffer optimization assays based on TLP activity performed by Brandon Iwaniec confirmed robust TLP activity in the classic assay buffer without DTT and BSA added, thus these components were removed from the CLIP binding buffer tested here. In my bead-binding assay results, we observe that FLAG-tagged DdiTLPs immunoprecipitate with higher yields than wild-types using anti-FLAG magnetic agarose after applying a 1000 mM NaCl/PBS wash step. We observe in the DdiTLP4 wild-type lanes that most wild-type protein is washed away in the flow-through. This demonstrates that the anti-FLAG agarose beads bind specifically to the FLAG epitope over other present peptides. Furthermore, in the FLAG-tagged TLP elution lanes, the band corresponding to the FLAG-DdiTLP is darker than potential other bands (other than the antibody heavy and light chains). For example, in the samples of untreated purified FLAG-tagged DdiTLPs, there are numerous lighter bands aside from the target protein band that represent impurities co-purified during metal ion affinity chromatography, but in the IP elution lanes, those bands are not visible or become even lighter (Figure 11). These observations further confirm that the anti-FLAG agarose is specifically binding with the FLAG epitope tag on DdiTLP3 and DdiTLP4, and also suggests that our salt wash protocol is effective in removing impurities present in the samples. These highly specific IPs with low nonspecific background give us confidence that our FLAG-tag system will work sufficiently for our CLIP studies.

Next, I optimized protocols for performing CLIP with formaldehyde as a crosslinking reagent. Standard formaldehyde crosslinking procedures in the literature use a formaldehyde concentration of 1%, incubating at room temperature for 10 minutes, and quenching with ~100mM glycine. I first sought to determine if these conditions will induce crosslinks in our in vitro system. I analyzed eluted samples with western blot and observe unique bands at high molecular weights in samples treated with formaldehyde, which may correspond to crosslinking. We see higher molecular weight bands and

smears in samples treated with formaldehyde that aren't present in the purified protein lane, which we attributed to the presence of large, crosslinked complexes (Figure 12A). The presence of these bands does not seem to depend on the presence of RNA, which is consistent with the knowledge that formaldehyde can produce protein-protein crosslinking. Furthermore, it appears that these bands and smearing are more intense in the elution lanes than flow throughs, suggesting that FLAG-tagged TLPs were abundant in the purified complexes. Thus, these results assured us that standard formaldehyde crosslinking conditions produce crosslinks in our CLIP system. Next, I compared proteinase K and heating as methods to extract crosslinked RNA from our IPs, utilizing radioactive labeling to observe IP RNA. Formaldehyde crosslinks are known to be reversible by heat, with crosslink half-life being as low as ~11.2 hours at 47°C (Jackson, 1978; Kennedy-Darling and Smith, 2014). Proteinase K is a broad-spectrum protease commonly used to digest enzymes, and is commonly used in CLIP studies to digest proteins and release crosslinked RNA. From numerous CLIP assays where I attempted both heat and proteinase K as IP RNA extraction methods, I more consistently was able to observe IP RNA extracted with proteinase K. Thus, out of caution I decided to use proteinase K to extract IP RNA in subsequent formaldehyde crosslinking optimization assays, however more trials of this comparison assay must be performed to draw sound conclusions on the true efficacy of proteinase K versus heat reversal. Since treatment with proteinase K involves incubation at a temperature that can range from 37-65°C, it is entirely possible that the heat conditions help to reverse the formaldehyde crosslinks while proteinase K digests the enzymes, making it a more robust method of IP RNA extraction than heat alone. Next, considering the propensity of formaldehyde to produce protein-protein crosslinks, I sought to determine if a formaldehyde concentration of less than 1% would produce crosslinks that can be readily observed with radiolabeling. Crosslinking is faintly evident at concentrations as low as 0.1 and 0.35%, shown by the appearance of very faint radiolabeled RNA bands on urea-PAGE, however, labeled RNA bands are most visible when treating with 1% formaldehyde, so out of caution I decided to use this concentration in

subsequent formaldehyde CLIP assays (Figure 12B). The lack of labeled RNA in negative control lanes is evidence that the RNA we do observe in experimental lanes is dependent on formaldehyde crosslinking it to protein. Our ability to resolve radioactively-labeled RNA extracted from CLIP pull-downs demonstrates that our protocols optimized for *in vitro* formaldehyde CLIP with FLAG-tagged DdiTLP4 are sufficient for capturing RNAs that interact with DdiTLPs during their RNA editing reactions. It is important to note the shortcomings of our use of this method, however. It was difficult to generate consistent, high-quality results; many trials did not detect radiolabeled RNAs despite the high sensitivity of radioactive labeling. These complications would make it difficult to reliably identify IP RNA with downstream quantitative methods. Further optimization with this method could potentially improve the results. Furthermore, formaldehyde is a highly toxic chemical; extreme caution must be taken when working with it in lab, and it requires special waste disposal. Thus, we look to performing *in vitro* UV CLIP with FLAG-tagged DdiTLP4 to compare this method with formaldehyde CLIP.

During the UV-CLIP protocol developed by Brandon Iwaniec for FLAG-tagged TLPs, rather than eliminating the crosslink to visualize IP'ed RNA like in formaldehyde CLIP, we maintain the crosslinks and visualize labeled protein-RNA complexes on SDS-PAGE protein gels. In this method, we radioactively label the 5' end of RNA crosslinked to the TLP, resolve the TLP-RNA complex on an SDS-PAGE gel, and transfer to a nitrocellulose membrane. Since nitrocellulose will only bind proteins, any radioactive label we detect on the membrane should in principle correspond to labeled RNA crosslinked to protein. I performed this UV-CLIP protocol with FLAG-tagged DdiTLP4 in numerous trials and obtained consistent results: the presence of an RNase A-dependent signal in the middle of the membrane, where we would expect the 37 kDa DdiTLP4 to run on an SDS-PAGE; the mass of the digested RNA fragment is assumed to be negligible compared to the protein size (Figure 13). The dependence of this band on RNase A concentration further confirms that we must be observing labeled RNA on the membrane, which must be crosslinked to protein on the nitrocellulose. A DdiTLP4 size marker would confirm that the labeled

RNA is indeed crosslinked to DdiTLP4, however, initial attempts at a Western blot of the nitrocellulose membrane with purified FLAG-tagged DdiTLP4 stock as a control were unsuccessful and must be reattempted. However, we do have enough evidence to conclude that our UV CLIP method can successfully and consistently immunoprecipitate TLP-RNA crosslinked complexes that can be readily observed with radiolabeling. Thus, we choose UV CLIP as our method to move forward with quantitative analysis of interacting RNAs. UV-CLIP is an increasingly used RNA immunoprecipitation method due to its ready ability to be coupled with high-throughput next-generation methods, an approach generally referred to as CLIP-seq. Modified CLIP methods such as PAR-CLIP and iCLIP (individual nucleotide resolution crosslinking and immunoprecipitation) are coupled with high-throughput RNA sequencing (RNA-seq) to identify interacting RNAs at their specific binding sites at single-nucleotide resolution. This is often done by generating dsDNA sequencing libraries through adapter ligation to IP RNAs and reverse transcription. However, many available CLIP-seq methods are technically challenging and produce high failure rates (Nostrand et al 2016). We are preparing to perform enhanced CLIP (eCLIP), developed by Nostrand et al (2016), a modified iCLIP method designed to successfully produce high-quality RNA-sequencing libraries. It involves the same general experimental workflow as our UV CLIP protocol but includes additional quality-control steps for preparing a high-quality dsDNA library. By using this method, we can quantify the RNAs that prefer to interact with DdiTLP4 *in vitro* and identify precisely where the enzyme binds its substrates at single nucleotide-resolution. This data will provide more insight to the possible biological substrates of this enzyme, specifically the class(es) of ncRNAs it prefers to act upon, and the state of RNA 5' ends it seeks to repair or act upon in some other way.

The formaldehyde and UV crosslinking protocols developed for in vitro CLIP with FLAG-tagged DdiTLP4 both present advantages and disadvantages for our purposes. While the formaldehyde method proved tricky to perfect *in vitro*, this method is still a powerful *in vivo* tool to purify crosslinked protein-protein complexes, which will serve to be useful for identifying the nuclease that likely complexes with

DdiTLP3 during its mitochondrial tRNA 5' editing reaction. UV light, while acting with a higher specificity, can only crosslink protein to RNA, which is a powerful method for our *in vitro* CLIP approach but can not give us insight to interacting proteins. Although we showed that formaldehyde also crosslinks DdiTLP4 to RNA, UV crosslinking produced more reliable, consistent results, and is used in the powerful eCLIP methodology, thus we choose UV crosslinking as our primary *in vitro* crosslinking reagent.

Chapter 3: *In vivo* CLIP with FLAG-tagged DdiTLP3 and DdiTLP4 in *D. discoideum*

While *in vitro* assays can be advantageous over *in vivo* work for numerous reasons, their results are not always completely reliable due to the altered chemical environment from biological conditions; sometimes, *in vitro* behavior of molecules does not exactly recapitulate its behavior *in vivo*. A relevant example of this is the observed activity of DdiTLP4 *in vitro*: it has the ability to repair 5' truncated tRNAs, yet there is no apparent need for that function in the cytosol, implying that DdiTLP4 could have functions beyond the 5' tRNA repair activity observed *in vitro*. Thus, it is advantageous to perform *in vivo* CLIP in tandem with our *in vitro* studies to identify RNAs and proteins that bind to DdiTLP3 and DdiTLP4 in their biological environments. First, we can identify RNAs crosslinked to FLAG-tagged DdiTLP4 *in vivo* with the eCLIP method and compare our results to the RNAs we identify *in vitro* to validate potential biological substrates. Secondly, performing formaldehyde CLIP with whole cell lysates allows co-purification of proteins that may interact with FLAG-tagged DdiTLP3 and DdiTLP4 in their biological environment, which we can identify with mass spectrometry and use to further understand each enzyme's mechanism.

Furthermore, performing *in vivo* CLIP is particularly convenient due to our access to its necessary experimental resources in our lab. For one, we have optimized *in vitro* CLIP protocols with FLAG-tagged TLPs, which we can easily adapt for performing *in vivo* CLIP. Secondly, our lab has experience working with *D. discoideum* strains. *Moreover, D. discoideum* is a genetically tractable organism and protocols for its growth and manipulation are well-established in the literature, allowing previous work in the lab to employ various *D. discoideum* genetic manipulation systems such as RNAi, DdiTLP2 deletion strains, and GFP expression systems, providing us a starting point for altering protein expression in other ways (Fey et al 2007; Eichinger and Rivero 2013; Long et al 2016). With these tools in hand, we can express FLAG-tagged versions of DdiTLP3 and DdiTLP4 in *D. discoideum* and perform *in vivo* CLIP to co-purify interacting proteins and RNA.

For this work we will grow AX-2 strains of *D. discoideum* that are engineered to grow axenically, or without bacteria present. Bacteria is a primary food source of this microorganism and cultures will double in as quickly as 4 hours when grown on bacteria (Fey et al 2007). We prefer to use pure cultures for simplification in our CLIP procedures; however, this means that they are easily contaminated with bacteria if not grown carefully. When growing *D. discoideum*, it is important to regularly split cultures to maintain healthy growth in the exponential phase; when cell density becomes high enough, cell growth enters stationary phase, when the organism begins developing due to starvation and altering the physiology of the culture (Fey et al 2007). For simplicity of cell culture methods and our CLIP studies, we maintain *D. discoideum* cells in their exponential growth phase. It is useful to find the doubling time of the culture during exponential growth so that a regular schedule for splitting cultures can be established to maintain healthy cells. Previous work in the lab found the approximate doubling time of AX-2 *D. discoideum* cell cultures, but since the saved strains had aged at least eight years since these experiments we thought it necessary to re-establish a growth curve and doubling time for reference in our experiments. By understanding the behavior of *D. discoideum* cell culture, we can successfully express FLAG-tagged DdiTLP3 and DdiTLP4 in this organism for *in vivo* CLIP. In this chapter we describe our assessment of the growth of *D. discoideum* and planned methods for expressing FLAG-tagged DdiTLP3 and DdiTLP4 i*n vivo*.

Results

7-day AX-2 D. discoideum growth curve calculates predictable culture doubling time

I sought to establish a 7-day growth curve for laboratory strains of *D. discoideum* that had been stored at -80 °C in order to calculate a doubling time of culture growth, which will be useful information for maintaining healthy cells to effectively perform *in vivo* CLIP. Cells were seeded at $5x10^4$ cells/mL and

grown for 7 days, being counted each day with a hematocytometer. The average cell density counted per day was plotted against time to calculate an average doubling time of 11.4 hours (Figure 14).

Adapting a system for expressing FLAG-tagged DdiTLP3 and DdiTLP4 in *D. discoideum*

Although we have not previously attempted to express FLAG-tagged TLP constructs in *D. discoideum* in our lab, we have the tools to do so. Previous work in the lab has taken advantage of the genetic tractability of this organism, manipulating gene expression with systems like RNAi, gene deletion via homologous recombination, and GFP expression. Thus, we plan to adapt previously engineered *D. discoideum* expression systems to express our FLAG-tagged DdiTLP3 and DdiTLP4. Dr. Yicheng Long successfully constructed pDXA cloning vectors, suitable for expressing proteins in *D. discoideum*, that contain C-terminal GFP-fused DdiTLP3 and and DdiTLP4 genes under control of an actin-15 promoter, which are inserted using double restriction cloning with HindIII and KpnI restriction sites. I plan to perform HindIII/Kpn1 restriction cloning to insert FLAG-tagged DdiTLP3 and DdiTLP4 genes into pDXA TLP-GFP vectors, which can be subsequently transformed into live *D. discoideum* AX-2 cells to express their proteins (Figure 15). We would include the stop codon of the DdiTLP gene to prevent expression of the GFP tag. However, the DdiTLP3 and DdiTLP4 genes both contain an internal HindIII restriction site, so internal-HindIII-deleted versions of these genes were created by Dr. Long. Thus, in order to successfully create expression vectors for FLAG-tagged DdiTLP3 and DdiTLP4 I must re-clone the FLAG tag onto the deleted-HindII versions of these genes as well as the HindIII and KpnI restriction sites for cloning into the pDXA vector. We are currently working to design primers and establish an appropriate workflow to properly clone FLAG-tagged DdiTLP3 and DdiTLP4 into *D. discoideum* expression vectors. When successful cloning is achieved, we can proceed to transforming our vectors into live cells to express FLAG-tagged DdiTLP3 and DdiTLP4 with actin-15 and perform *in vivo* CLIP.

Discussion

In this chapter, we present preliminary work towards performing CLIP *in vivo* in *D. discoideum.* I first grew saved strains of AX-2 *D. discoideum* to generate a 7-day growth curve and calculate an average doubling time of cell growth. The doubling time I found, 11.4 hours, is consistent with the typical doubling time of axenically growing *D. discoideum,* leading us to conclude that the eight-year-old strains still proliferate readily in axenic conditions, and that our HL5 media and antibiotic concentrations (optimized by Maria Abad and Yicheng Long) are suitable for the growth of these cells (Fey et al 2007). Establishing a predictable doubling time for our AX2 strains will help us maintain the health of live cultures so that we can properly express FLAG-tagged DdiTLP3 and DdiTLP4 from pDXA vectors. I am currently working on cloning pDXA vectors containing FLAG-tagged DdiTLP3 and DdiTLP4 using a previously created DdiTLP-GFP expression vector. We anticipate that our cloning workflow may be tricky because of the need for a lengthy primer on the terminus that includes the FLAG tag as well as the necessary restriction site, which could impact complete annealing of the primer during PCR. After confirming the insertion of the FLAG-TLP gene into the pDXA-GFP vector with Sanger sequencing, we can transform it into *D. discoideum* with electroporation, express the FLAG-TLP gene by induction of the actin-15 promoter, and immunoprecipitate FLAG-tagged DdiTLP3 and DdiTLP4 crosslinked to interacting molecules from their biological environment.

Chapter 4: Conclusion and Future Studies

Throughout this work we sought to further characterize the 3'-5' polymerase functions of DdiTLP3 and DdiTLP4 by identifying molecules that physically interact with them during their reactions. To do this, I successfully cloned, expressed, and purified FLAG epitope-tagged versions of DdiTLP3 and DdiTLP4 in high enough yields to be used in i*n vitro* immunoprecipitation experiments to co-purify interacting RNAs or proteins. I created N-terminal and C-terminal FLAG tagged DdiTLP4, and C-terminal FLAG tagged DdiTLP3. We then sought to determine if placing a FLAG tag on a terminus of DdiTLP3 and DdiTLP4 would impact their 3'-5' nucleotide addition activity to ensure that we could perform successful and accurate immunoprecipitations, and found with qualitative and kinetic activity assays that FLAG-tagged DdiTLP3 and DdiTLP4 retain comparable activity to wild-types. Previous attempts at fusing a FLAG tag to DdiTLP2 were unsuccessful, so this is the first evidence of a Thg1 family enzyme being amenable to epitope tagging.

With confirmation that the FLAG tag would most likely not impact the binding of interacting molecules, we established protocols for *in vitro* crosslinking and immunoprecipitation (CLIP) with FLAG-tagged TLPs, comparing formaldehyde and UV light as crosslinking reagents to identify potential RNA substrates of DdiTLP4 from total RNA isolated from *D. discoideum.* In our *in vitro* studies we were able to utilize radioactive labeling to observe RNAs crosslinked to FLAG-tagged DdiTLP4 with both tested crosslinking reagents. However, in our studies with formaldehyde as a crosslinking reagent, we were not as able to reproducibly immunoprecipitate and visualize substantial amounts of RNA. This was possibly caused by the added experimental complexity with reversing protein-RNA crosslinks before radiolabeling, or the lower specificity of formaldehyde as a crosslinker, which likely produced protein-protein crosslinks that may have interfered with substantial creation of enough protein-RNA crosslinks for detection. Conversely, with the UV CLIP protocol described here, which preserve the irreversible

protein-RNA crosslink induced by UV, we were able to consistently observe radiolabeled RNA crosslinked to protein. Additionally, advanced next-generation sequencing methods for identifying IP RNAs have been established downstream from modern UV crosslinking protocols, like the eCLIP we plan to perform. Furthermore, the high specificity of UV light as a crosslinker is beneficial for our *in vitro* purposes since the only crosslinks we obtain are between protein and RNA. Thus, we decided to focus on the UV CLIP protocol as our primary pathway to identifying potential RNA substrates of DdiTLP4 *in vitro*. More optimization work with the *in vitro* formaldehyde CLIP method may be needed to make it suitable for pairing with high-quality RNA sequencing methods like those we plan to perform with UV CLIP. It would be interesting to perform formaldehyde CLIP with the same experimental workflow as that which we use for UV CLIP, analyzing protein-RNA complexes rather than RNA alone; however, the lengthy heating steps associated with preparing crosslinked complexes for analysis would likely reverse the formaldehyde-induced crosslinks, demonstrating another utility of UV CLIP over formaldehyde CLIP *in vitro*. Either way, our *in vitro* CLIP studies performed so far using both crosslinking reagents were largely qualitative, showing that the protocols we developed were suitable for our experimental goals, but will require additional steps to obtain the sequences of any crosslinked RNA species. Our planned studies with eCLIP for applying next-generation sequencing to our *in vitro* CLIP pull-downs will provide the necessary quantitative data to draw conclusions about potential DdiTLP4 substrates.

While we prepare to perform eCLIP to sequence *in vitro* RNA substrates of DdiTLP4, we prepare to perform *in vivo* CLIP experiments to identify interacting molecules in a true biological context. Although *in vivo* CLIP would require modification of various steps of our *in vitro* CLIP protocol, most of the available CLIP protocols we drew inspiration from are performed *in vivo*, so we anticipate that we can revisit these literature sources and easily adapt an *in vivo* CLIP protocol with our FLAG-tagged DdiTLPs. This would involve treating live cells with the crosslinking reagent and subsequently lysing the cells to release FLAG-TLP complexes for immunoprecipitation. Co-purifying and identifying proteins and

RNAs that interact with DdiTLP3 and DdiTLP4 in their true biological environment rather than *in vitro* conditions may be a more robust method for fully understanding the molecular machineries that facilitate their 5' RNA processing activities, including interacting proteins that aren't accounted for in our *in vitro* RNA CLIP. Crosslinking *in vivo* allows us to “freeze” active reaction complexes in real time, which we can analyze to further understand the complete 5' RNA processing pathways DdiTLP3 and DdiTLP4 are players in. For example, we can potentially identify upstream components that prepare substrates for processing by a 3'-5' polymerase, like the nuclease that acts in the 5' tRNA editing reaction with DdiTLP3. The identity of this nuclease in the mitochondrial tRNA 5' editing reaction has been long sought after, and our identifying it in *D. discoideum* with *in vivo* CLIP will aid in identifying its homologs in other species where 5' editing occurs. Identifying upstream proteins, which could include a similarly acting nuclease, that may interact with DdiTLP4 will be useful for understanding the activities that renders an RNA a substrate for DdiTLP4 and thus the need for 5' RNA processing in the cytoplasm. This could help confirm the proposed model that DdiTLP4 processes Class I ncRNAs involved in *D. discoideum* development, as well as help us understand why potential DdiTLP4 substrates sequenced from DdiTLP4 depletion strains have varying distributions of nucleotides at their 5' ends rather than a discrete terminus to act on. Furthermore, by understanding the need for cytoplasmic RNA 5' processing in *D. discoideum*, we can apply our knowledge to identifying the roles of cytoplasmic TLPs expressed in other organisms, as well as understand how the broad cytoplasmic ncRNA editing by DdiTLP4 and homologs evolved into the specific G_{-1} function performed by Thg1 in higher eukaryotes. Thus, it will be very useful to supplement *in vitro* RNA CLIP studies with *in vivo* CLIP to obtain a more complete picture of the potential ncRNA processing pathways DdiTLP4 is implicated in.

While Thg1 performs a very specific function in eukaryotic cells, previous investigations into the catalytic activities of the evolutionarily ancient TLPs, especially of the four distinct ones discovered in the eukaryotic *D. discoideum*, suggests to us that these Thg1 family members are capable of broader

RNA processing activities. Thus, they could be implicated in previously unknown or incompletely understood RNA quality control pathways, like the mitochondrial tRNA 5' editing for which DdiTLP3 was found to play the role of the elusive 5' repair enzyme. It is likely that these RNA processing pathways involve a host of molecular machinery that facilitate and regulate each step, thus identifying molecules that co-purify with an epitope-tagged TLP will provide more insight into their tRNA and ncRNA processing activities. This work represents the first attempt in our lab to perform crosslinking and immunoprecipitation (CLIP) with an epitope-tagged Thg1 family member. By showing that DdiTLP3 and DdiTLP4 are amenable to epitope tagging, we can harness this technology and apply it to other Thg1 family members to further understand their tRNA (or other ncRNA) processing activities, allowing us to composite a more complete portfolio of the capabilities of 3'-5' polymerases, as well as establish the evolutionary descent of TLP functions to that of eukaryotic Thg1.

Materials and Methods

Cloning, expression, and purification of FLAG-tagged DdiTLP3 and DdiTLP4 constructs

In previous work in the lab, the DdiTLP3 and DdiTLP4 genes were cloned into an AVA421 vector containing an N-terminal His_6 tag, an ampicillin resistance gene (beta-lactamase) and an IPTG-inducible T7 promoter (Abad et al. 2011). Oligonucleotide primers were designed to insert the FLAG sequence at the desired terminus of each protein gene; due to the length of the FLAG nucleic acid sequence, forward and reverse primers were created with one half of the insertion sequence on each of their 5' ends. Polymerase chain reaction (PCR) with Phusion site-directed mutagenesis with our primers was used to insert the FLAG sequence into the DdiTLP-AVA421 plasmids. Each reaction was treated with 20U DpnI restriction enzyme. Following re-circularization of the product DNA with T4 DNA ligase, the plasmids were transformed using heat-shock into competent BL21 *E. coli* cells and subsequently purified with a miniprep. Sanger sequencing confirmed that the FLAG sequence was successfully inserted into each DdiTLP gene. The purified mutant plasmids were transformed via heat shock into *E. coli* Rosetta pLysS competent cells, cultures grown in LB media supplemented with 100ug/mL ampicillin, and the FLAG-tagged TLP genes were overexpressed with 1mM IPTG when each culture reached OD_{600} = [0.4]. Following an induction incubation time at temperatures previously optimized for each TLP gene (4 hrs at 37°C for DdiTLP4 and 18°C O/N for DdiTLP3), cells were harvested by centrifugation and lysed by sonication. The desired proteins were subsequently purified using TALON immobilized cobalt-ion affinity chromatography and dialyzed into 50% glycerol for storage at -20*C. Protein concentrations were measured with the BioRad Bradford protein assay (Bradford, 1976) and the purity of proteins assessed with SDS-PAGE as judged by the intensity of Coomassie-stained protein bands, 6 uG of each loaded into each lane to normalize.

Immunoblotting of FLAG-tagged DdiTLP3 and DdiTLP4

We utilized the western blot to specifically detect the presence of a FLAG epitope tag on each DdiTLP construct with wild-types used as controls. 1ug of each enzyme was twice serially diluted 100-fold, and each titration run on an SDS-PAGE gel before electrophoretic transfer to a nitrocellulose membrane in a 20% methanol transfer buffer. The membranes soaked in 10x PBS supplemented with 0.05% NaCN overnight. The next day 3% milk was used to block the membrane for 1 hour. Anti-FLAG antibody produced in mouse was diluted to 4ug/mL in 3% milk and incubated with the membrane for 2 hours. 0.8 mg/mL anti-mouse igG antibody with conjugated horseradish peroxidase (HRP) was diluted 1/10,000 into 3% milk and incubated with the membrane for 1 hour. The membrane was stringently washed with Milli-Q water between each incubation step. Finally, the membrane is developed with an ECL kit, rinsing with a 1:1 solution of luminol and peroxide to react with the HRP and produce light that can be detected with a film developer.

Phosphatase protection assays to determine the catalytic efficiency of FLAG-tagged DdiTLP3 and DdiTLP4

Phosphatase protection assays (PPAs) were carried out to compare the catalytic efficiencies of FLAG-tagged DdiTLPs with their wild-type counterparts. *D. discoideum* mt-tRNA^{Trp} with a G_{+1} 5' start was chosen as the substrate for this assay. The substrate was *in vitro* transcribed with T7 RNA polymerase from a previously constructed vector containing the tRNA sequence under control of a T7 promoter. The tRNA was treated with calf-intestinal phosphatase (CIP), then with γ-[32 P]ATP and polynucleotide kinase (PNK) to radioactively label its 5' end. Each enzyme was incubated with the labeled tRNA substrate and 1mM of the nucleotide to be repaired with (ATP) at 2000 cpm per reaction in buffer conditions previously determined to be suitable for TLP activity (25mM HEPES, 10mM MgCl2, 100mM DTT, 125mM NaCl, 0.2 mg/mL BSA). For end-point assays, titrations of each enzyme are allowed to react

for 2 hours. For time-point assays, aliquots of reactions with were quenched at time points over the course of 2-2.5 hours. RNase T1 with 10mM EDTA, 100mM NaoAC, and 2ug yeast tRNA quenched the reactions. Each reaction is treated with 0.5U CIP and the products resolved with TLC in a 55:35:10 n-propanol: NH_4OH: H_2O solvent system. TLC plates are exposed to a light screen to be scanned on a Typhoon phosphorimager. Time-point assays were quantified with the ImageQuant software, and the data was plotted and fitted to a single-exponential rate equation to calculate the reaction rate k_{obs} using the Kaleidagraph statistics software:

$$[P]_t = [P_{max}](1-e^{-kt})$$

Where $[P]_t$ is the % product formation at any time t, and $[P]_{max}$ is the maximum observed % product formation. The mean k_{obs} was calculated for each enzyme from triplicate (n=3) data sets, although only two replicates (n=2) were performed for C-terminal FLAG DdiTLP4.

Bead binding assays to test efficacy of IP with FLAG tags

We use the Pierce Anti-DYKDDDDK magnetic agarose beads by ThermoFisher to perform our IPs, which have a binding capacity of 3.2 mg /mL. 120uL of beads were washed with 500uL of 1X PBS three times over, nutating for at least 5 minutes between each wash. Washes were separated from beads with a Cortex Biochem magnetic stand. 40ug of protein (to max binding capacity) was incubated with the beads overnight at 4*C, nutating in 500uL of binding buffer (described previously). The flow-through was separated with the magnetic stand, and then beads were washed three times with 1x PBS, nutating for 5 minutes between each wash. Fractions of each wash step were saved for SDS-PAGE analysis. To elute the FLAG-proteins, beads were resuspended in 100uL 1X SDS-PAGE sample buffer and boiled at 95-100*C for 5-10 minutes, and then separated with the magnetic stand. The flow-through samples were treated with 72% w/v TCA to 10% of the flow-through solution, incubated for 30 minutes, and

resuspended in SDS-PAGE sample buffer supplemented with 5uL of 2M tris base to be boiled like the elution samples. All samples were analyzed with SDS-PAGE.

***In vitro* CLIP with formaldehyde as a crosslinking reagent**

D. discoideum total RNA was previously extracted from live cultures with TRIzol reagent. 40-50ug of *D. discoideum* total RNA was incubated with 150ug C-terminal FLAG DdiTLP4 in 100uL binding buffer (previously described) at room temperature for 30 minutes to form RNA-protein complexes. Formaldehyde, 37% w/v, was added to each reaction to 1% concentration, and incubated at room temperature for 10 minutes. Glycine was added to a final concentration of 100mM to quench each crosslinking reaction. Samples were transferred to washed anti-FLAG agarose beads (bead wash steps previously described) and nutated overnight at 4*C. Beads were washed as previously described. Samples from each wash step were collected for SDS-PAGE analysis. For SDS-PAGE and western blot analysis, DdiTLP-RNA complexes were eluted from the beads in SDS-PAGE loading buffer boiled at 95-100*C for no longer than 5 minutes before being run on a gel. We choose 5 minutes so that complexes are eluted off the beads while the formaldehyde crosslinks, which are reversible by heat, are maintained. Western blot analysis was performed with anti-FLAG antibodies as previously described. For urea-PAGE analysis, we elute IP RNA from bound proteins by digesting the protein with 10U Invitrogen Proteinase K and then perform a PCA extraction. Extracted RNA is treated with 1 uL Quick CIP (NEB), and 2 uL T4 polynucleotide kinase (NEB) with at least 0.01 uL of γ-[32 P]ATP; CIP/Kinase reactions are performed as described in product manuals. Labeling reactions are quenched with urea-PAGE loading dye, run on a 10% acrylamide 4M urea gel, and exposed to a light screen overnight to be resolved with a Typhoon phosphorimager.

***In vitro* CLIP with UV light as a crosslinking reagent**

D. discoideum total RNA was previously extracted from live cultures with TRIzol reagent. 30-50ug of *D. discoideum* total RNA was incubated with 100-150ug C-terminal FLAG DdiTLP4 in 200-400uL binding buffer (previously described) at room temperature for 30 minutes to form RNA-bound complexes. Equal volumes of the solution were divided into the desired number of wells on a 96-well plate corresponding to the desired amount of reaction conditions. The plate was positioned ~3cm below the UV lamp in a Stratalinker and exposed to ~2.5-4.2 total Joules of energy at ~2.8-3.2 J/cm^2. Crosslinked samples were loaded onto 100-150uL washed anti-FLAG agarose beads and nutated overnight, subsequently washing 3x with 500 mM NaCl wash buffer. Samples were treated with 10uL RNase A at concentrations ranging from 10ug/mL to 10 mg/mL in our binding buffer and incubated at 37*C for 5 minutes. Next, samples are treated with 2 uL of NEB Quick CIP for 60 minutes at 37*C in 1x NEB rCutSmart buffer followed by CIP inactivation at 85°C. Crosslinked RNA is radioactively labeled with the addition of T4 PNK and γ-[32 P]ATP, incubated at 37°C for 60 minutes. It is important to note that between each step, beads are washed with NaCl wash buffer 3x. Complexes are eluted from beads in 50uL SDS-PAGE sample buffer boiled at 90-100*C for 10 minutes and separated with the magnetic stand. Samples are run on an SDS-PAGE gel and then transferred to a nitrocellulose membrane with a 20% methanol tris-base/glycine transfer buffer at 100 volts for at least 60 minutes. The membrane is stained with Ponceau to visualize transferred protein, and then exposed overnight to a light screen to be scanned on a Typhoon phosphorimager.

Growth of *Dictyostelium discoideum*

Protocols for growing *D. discoideum* were previously established in the lab, mostly derived from those published by Fey et al. (2017) and Eichinger and Rivero (2013). AX-2 strains of *D. discoideum* were previously obtained from the Dicty Stock Center at Northwestern University and save strains are stored in our lab at -80°C in 10% DMSO. Strains were quickly thawed in a 37*C water bath before being diluted

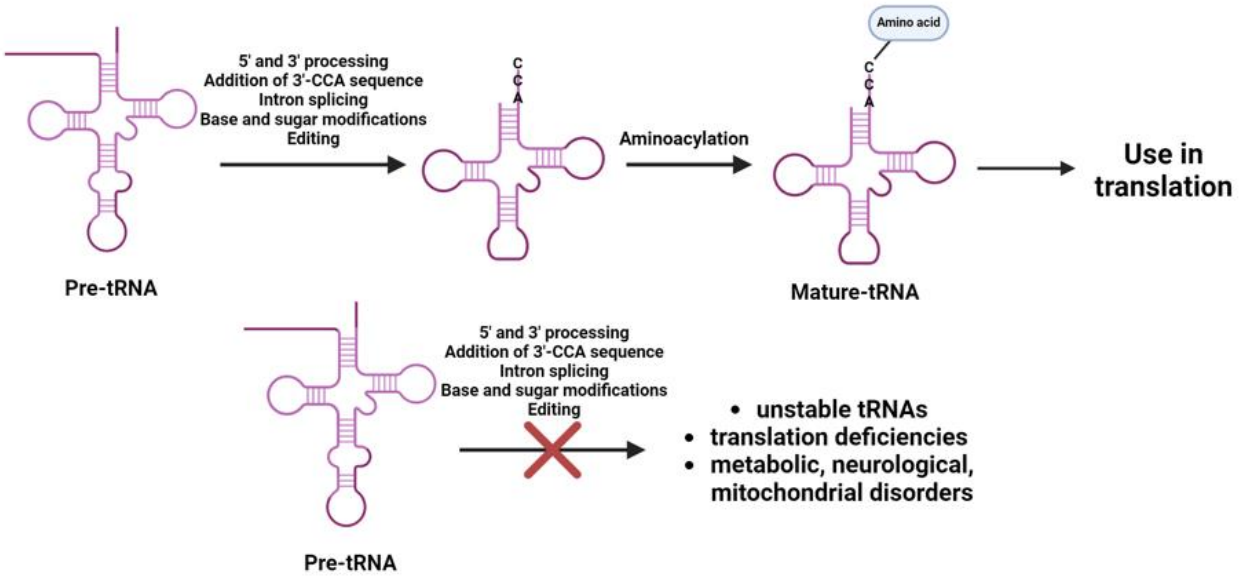
1/10 into HL5 media without antibiotics. All work with cells outside of the incubator was done in sterile conditions on the bench top as AX-2 strains are easily contaminated with bacteria and yeast. After the cells recovered at 22°C for 60 minutes the media was changed by separating the cells with centrifugation and then allowed to incubate at 22°C overnight, and from then on media was supplemented with 50 ug/mL ampicillin and 40 ug/mL streptomycin. Cells incubated at 22°C and were counted each day with a hemacytometer, with plates being split as necessary to keep cell density below 4×10^6. To produce the growth curve, thawed cells were grown for several generations before being seeded at 5×10^4 cells/mL and counted each day for 7 days, splitting cultures as necessary. Growth rates and doubling time are calculated with the following equations:

$$gr = \frac{\ln(\frac{N_t}{N_0})}{t} \qquad doubling\ time = \frac{\ln(2)}{gr}$$

After splitting plates, the average cell density from all plates is used for the growth curve, accounting for the dilution factor.

Figures

A)



B)

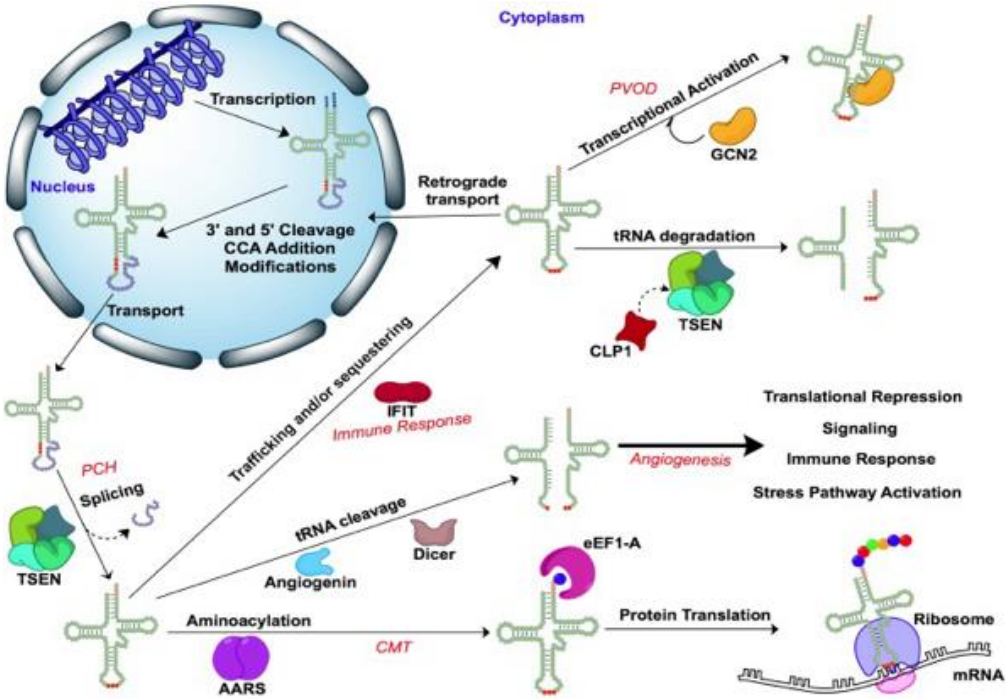


Figure 1: Post-transcriptional tRNA processing is vital to cellular function. A) General chronology of post-transcriptional modifications to a tRNA species. B) Birds-eye view of some post-transcriptional tRNA interactions and pathways in a cell. From Abbott et al. 2014.

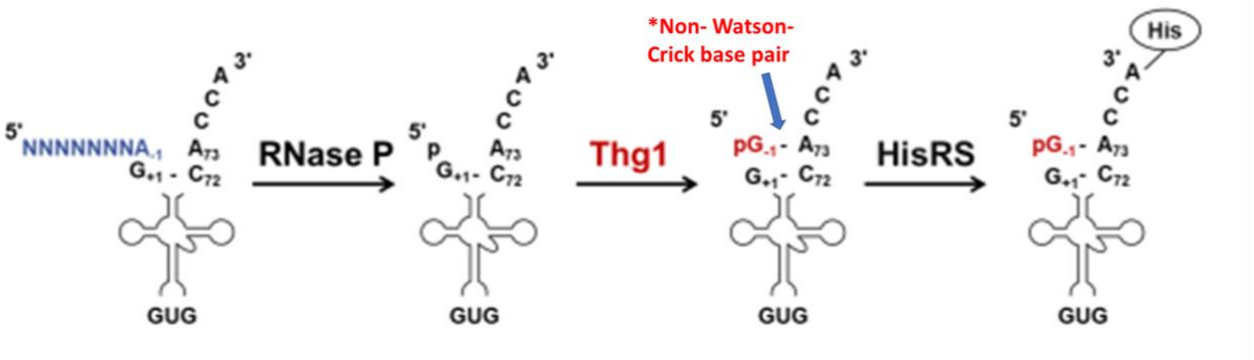


Figure 2: Thg1 mechanism of action. In eukaryotes, $tRNA^{His}$ guanylyltransferase (Thg1) catalyzes the non-canonical 5' addition of a single G-1 to tRNAHis, a universally conserved identity element for recognition by histidyl-tRNA synthetase (HisRS). Adapted from Jackman et al. (2012) and Dodbele et al (2019)

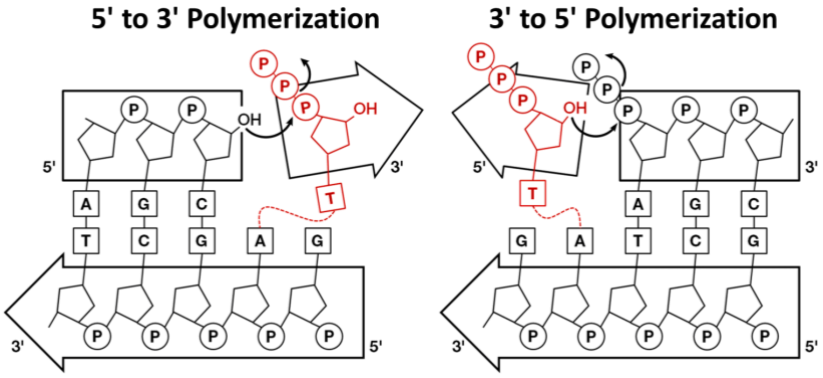


Figure 3: 5-3' polymerization and 3'-5' polymerization mechanisms are chemically equivalent. Adapted from Dodbele et al.

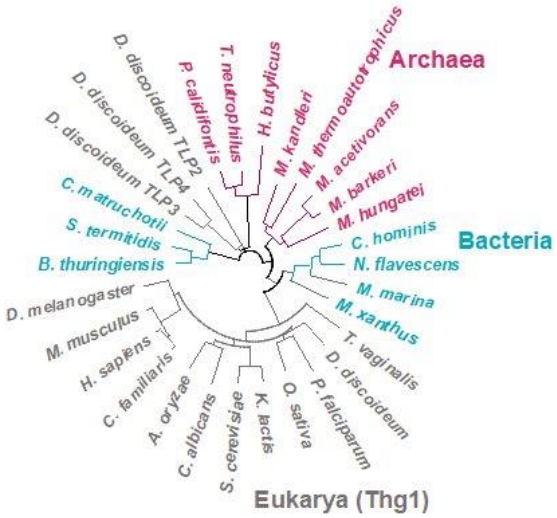
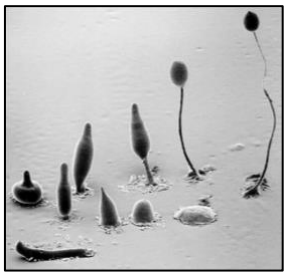


Figure 4: Phylogeny of the Thg1/TLP enzyme family. TLPs are found in all domains of life, while Thg1 only is retained in eukaryotes.

A)



Copyright, M.J. Grimson & R.L. Blanton

B)

Ddi Thg1/TLP	Cellular localization	Biological function	Essential?
DdiThg1	Cytosol	G_{-1} addn' to 5' end of $tRNA^{His}$	✓
DdiTLP2	Mitochondria	G_{-1} addn' to 5' end of mt-$tRNA^{His}$	Growth defect if deleted
DdiTLP3	Mitochondria	Mitochondrial tRNA 5' editing	✓
DdiTLP4	Cytosol	Non-coding RNA processing and/or repair function?	✓

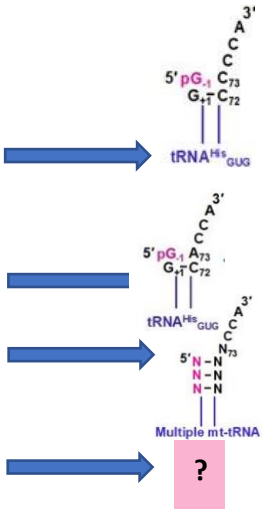


Figure 5: *D. discoideum* is a useful model organism for studying the Thg1/TLP enzyme family. A) *D. discoideum* development, from M.J. Grimson and R.L. Blanton. B) Summary of the functions of each Thg1/TLP enzyme found in *D. discoideum.*

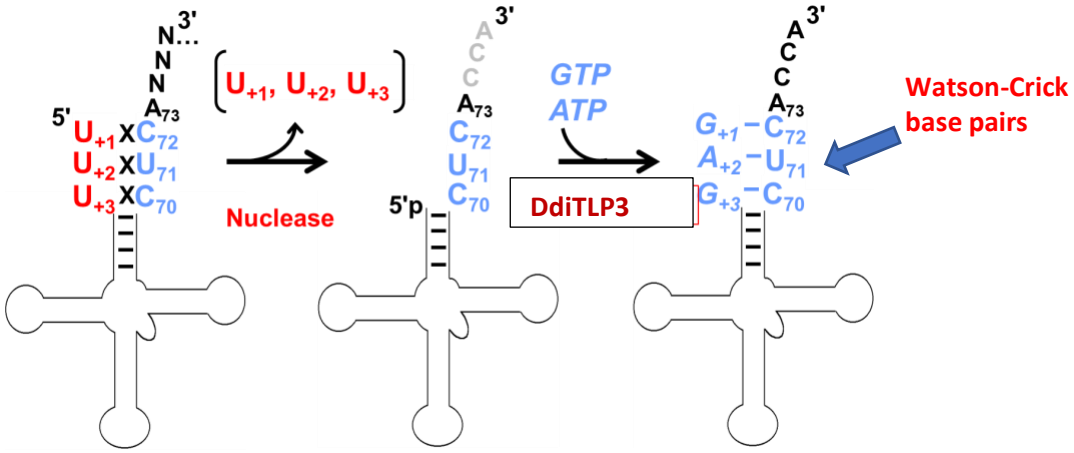


Figure 6: Mechanism for mitochondrial tRNA 5' editing, for which DdiTLP3 is the first known candidate of the 3'-5' repair enzyme. Consistent with TLP catalytic activity, DdiTLP3 repairs nucleotides in a Watson-Crick base-paired manner.

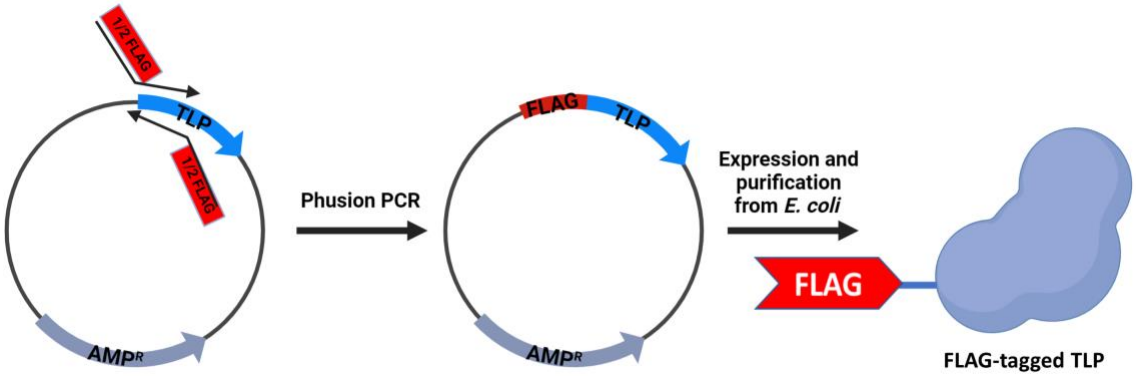
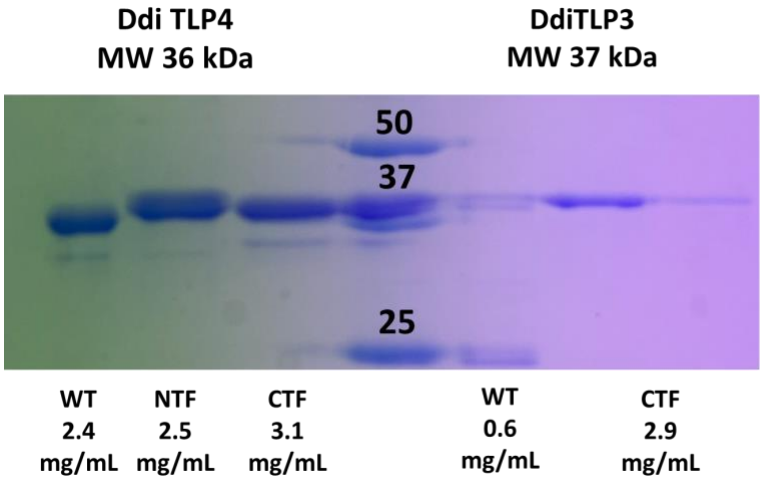


Figure 7: Workflow for cloning a FLAG epitope tag to the terminus of a DdiTLP protein, using an AVA421 vector containing the TLP gene and suitable PCR primers to insert the FLAG sequence.

A)



B)

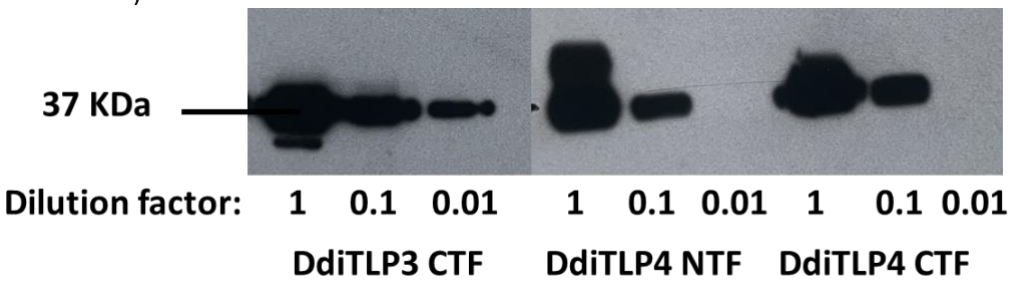
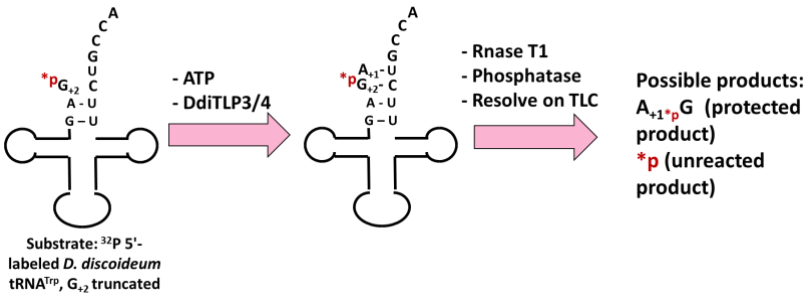
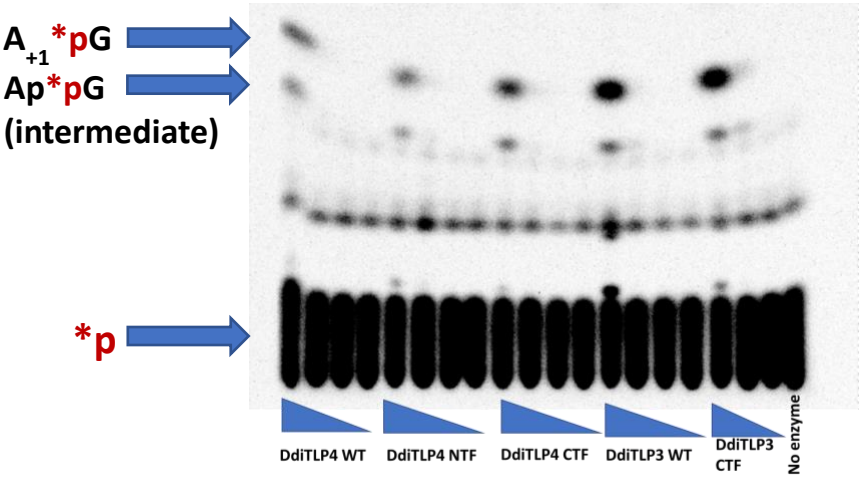


Figure 8: Purification of FLAG-tagged DdiTLP3 and DdiTLP4. A) SDS-PAGE analysis of purified DdiTLP3 and DdiTLP4 wild-types and FLAG-tagged constructs. B) Western blot of FLAG-tagged DdiTLP constructs with anti-FLAG primary antibody. MW=Molecular Weight; NTF=N-terminal FLAG; CTF=C-terminal FLAG

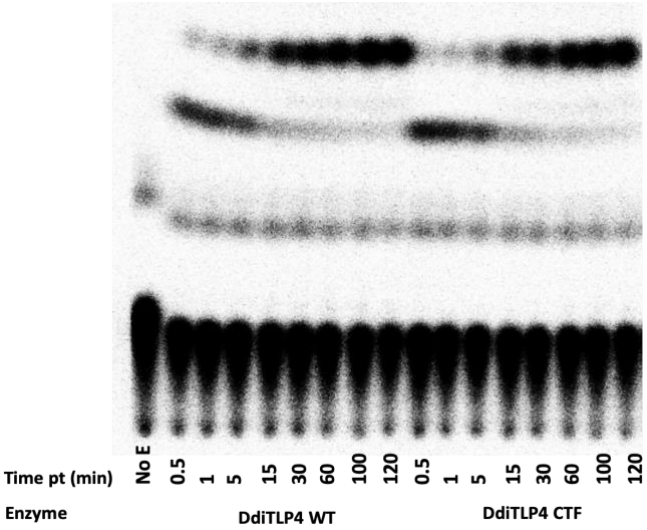
A)



B)



C)



D)

Enzyme	k_{obs} (min^{-1})
WT DdiTLP4	0.041 ± .013
CT-FLAG DdiTLP4	0.085 ± .040
WT DdiTLP3	0.038 ± .009
CT-FLAG DdiTLP3	0.024 ± .011

Figure 9: Phosphatase protection assay with FLAG-tagged DdiTLP3 and DdiTLP4 versus wild-types. A) Schematic of the workflow for the assay. B) End-point assay run for 2 hours with serial dilutions of enzyme. C) One trial of a time-point assay with DdiTLP4 CTF versus wild-type. D) Single-turnover kinetic parameters for each enzyme determined from time-point assay data.

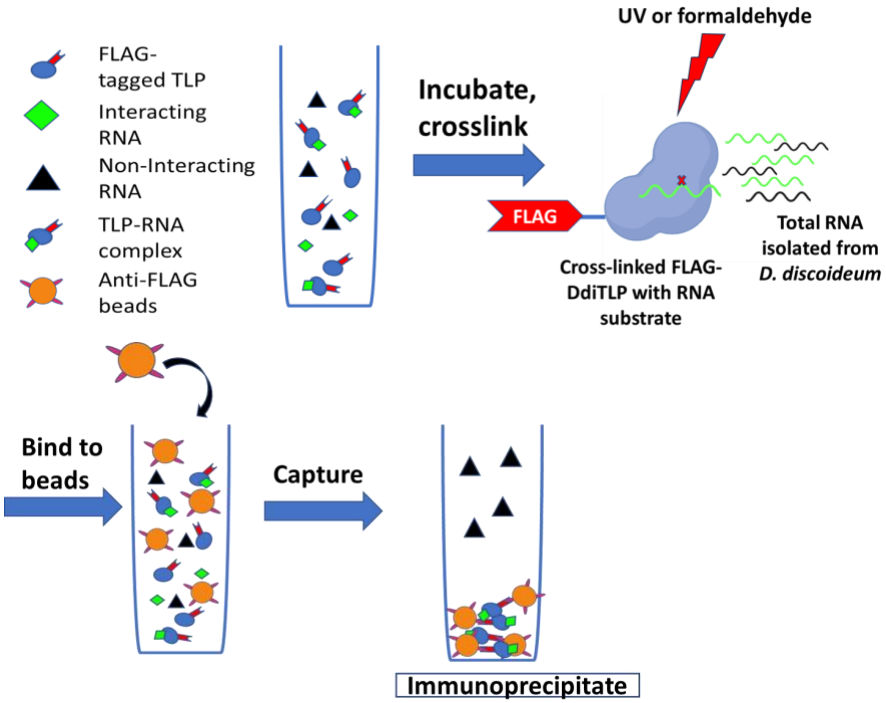
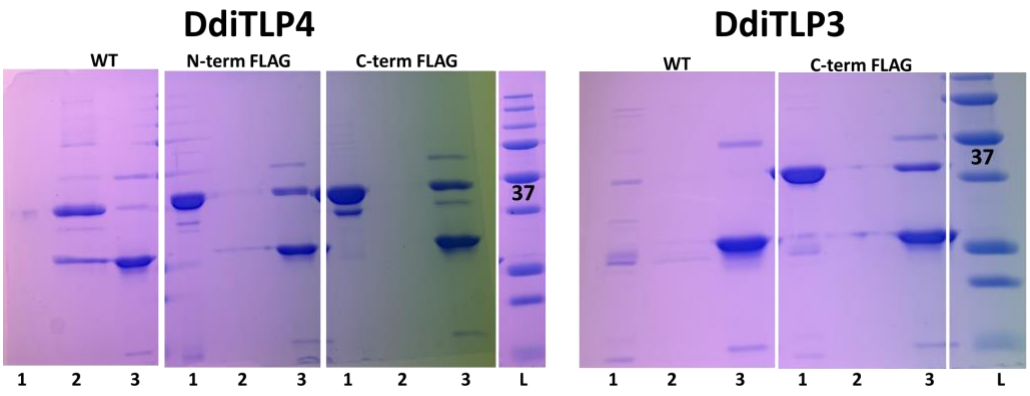


Figure 10: Schematic for performing *in vitro* crosslinking and immunoprecipitation (CLIP) with FLAG-tagged DdiTLPs.



Legend

Protein	Approx. size (kDa)
DdiTLP3	37
DdiTLP4	36

1	Purified protein
2	Flow-through
3	Eluted protein from anti-FLAG beads
L	Molecular weight ladder (kDa)

Figure 11: Bead-binding assay without RNA or crosslink reagent present to confirm efficacy of IP protocol FLAG specificity

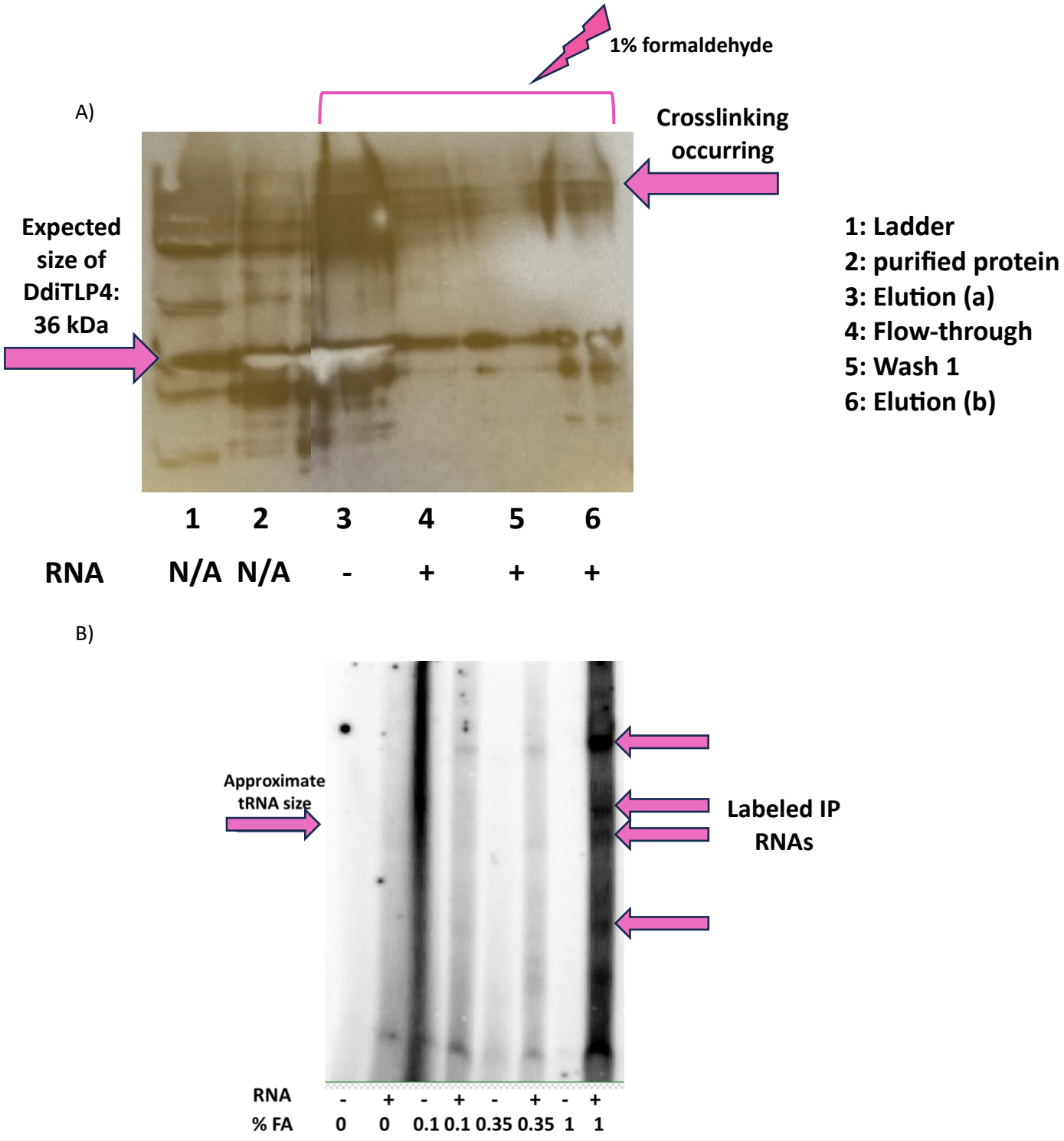


Figure 12: Formaldehyde can effectively crosslink *in vitro.* A) Western blot of formaldehyde CLIP samples compared to untreated purified DdiTLP4. Elution sample was split into lanes 3 and 6. Expected size of DdiTLP4 is 36 kDa. B) Urea-PAGE of radiolabeled RNA extracted from formaldehyde CLIPs to optimize formaldehyde concentration.

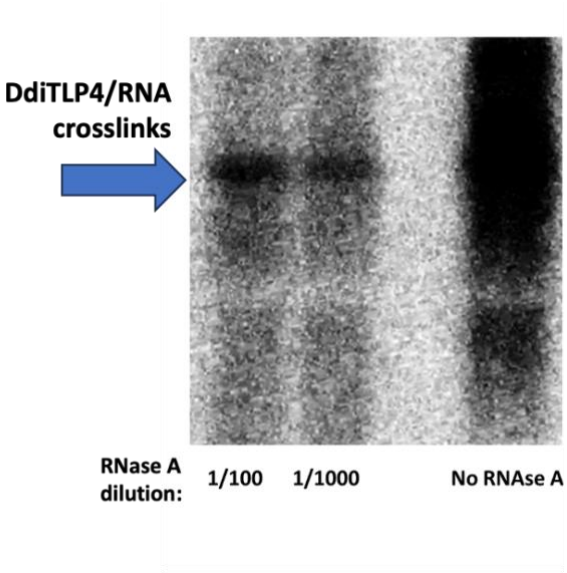


Figure 13: Nitrocellulose membrane transferred from SDS-PAGE of protein UV-crosslinked to radiolabeled RNA, showing that UV effectively crosslinks DdiTLP4 to RNA *in vitro.*

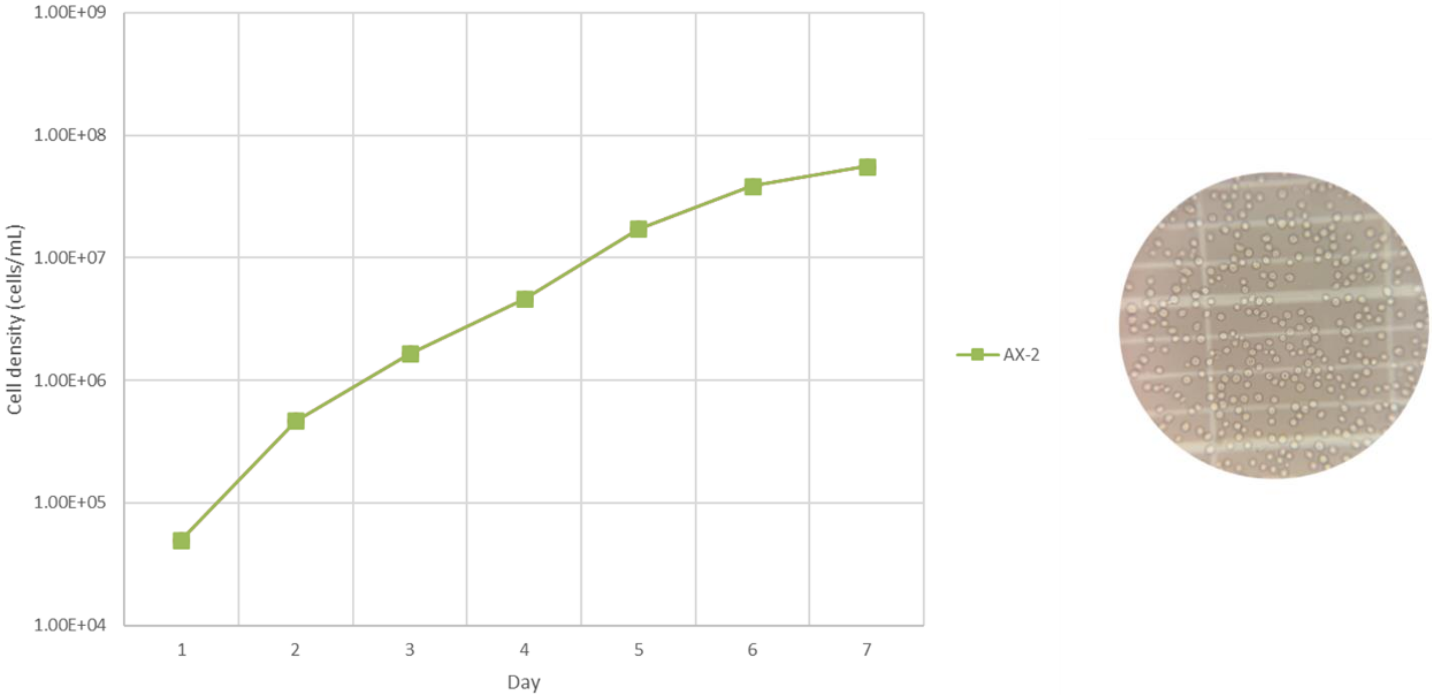


Figure 14: Growth curve of axenically-growing *D. discoideum* cell culture.

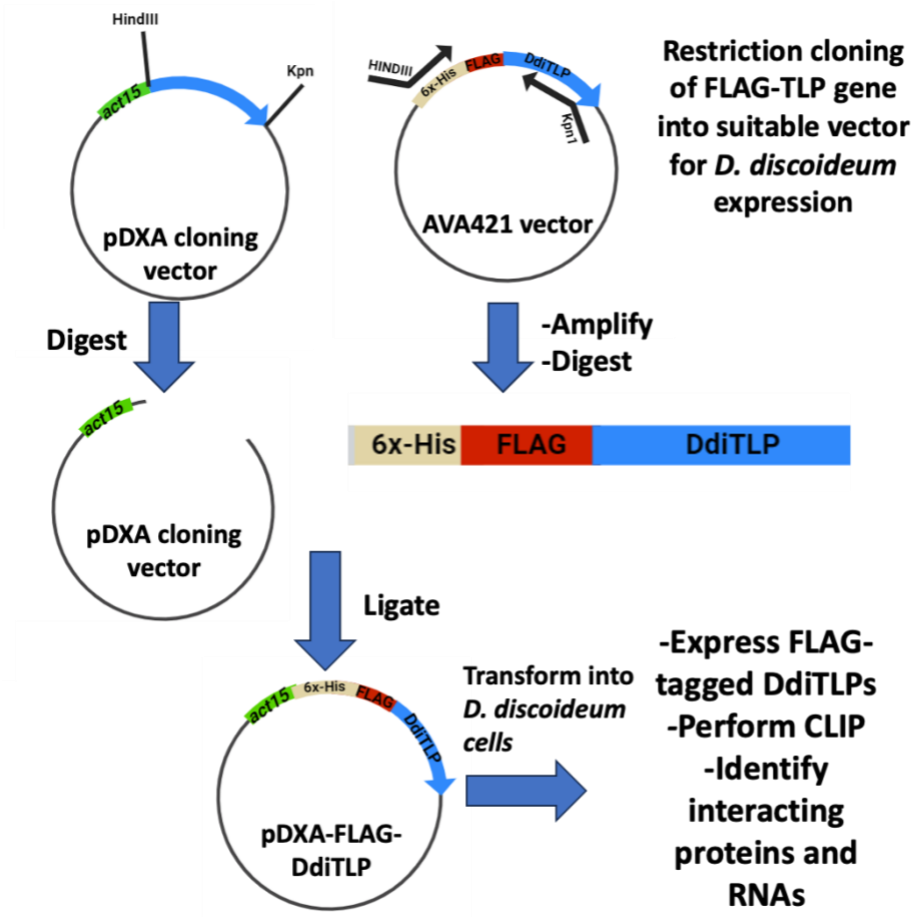


Figure 15: Schematic for procedure to clone vectors suitable for expressing FLAG-tagged DdiTLPs *in vivo* to perform *in vivo* CLIP

References

1. Abad M.G., Rao B.S., Jackman J.E. Template-dependent 3’-5’ nucleotide addition is a shared feature of tRNAHis guanylyltransferase enzymes from multiple domains of life. Proc. Natl. Acad. Sci. U.S.A. 2010;107:674–679.
2. Abad M.G., Long Y., Willcox A., Gott J.M., Gray M.W., Jackman J.E. A role for tRNA(His) guanylyltransferase (Thg1)-like proteins from Dictyostelium discoideum in mitochondrial 5’-tRNA editing. RNA. 2011;17:613–623
3. Abbott, J. A., Francklyn, C. S., & Robey-Bond, S. M. (2014). Transfer RNA and human disease. Frontiers in Genetics, 5. https://doi.org/10.3389/fgene.2014.00158
4. Bradford MM (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72. 248–254.
5. Brizzard B. Epitope tagging. Biotechniques. 2008 Apr;44(5):693-5. doi: 10.2144/000112841. PMID: 18474046.
6. Bullerwell, C. E., & Gray, M. W. (2005). In vitro characterization of a trna editing activity in the mitochondria of Spizellomyces punctatus, a chytridiomycete fungus. Journal of Biological Chemistry, 280(4), 2463–2470. https://doi.org/10.1074/jbc.m411273200
7. Burkard U, Söll D 1988. The 5’-terminal guanylate of chloroplast histidine tRNA is encoded in its gene. J Biol Chem 263: 9578–9581
8. Burkard U, Willis I, Söll D 1988. Processing of histidine transfer RNA precursors. Abnormal cleavage site for RNase P. J Biol Chem 263: 2447–2451
9. Cooley L, Appel B, Soll D 1982. Post-transcriptional nucleotide addition is responsible for the formation of the 5' terminus of histidine tRNA. Proc Natl Acad Sci 79: 6475–6479

10. Dodbele, S. (2019). Characterization of the function and mechanism of an orphan 3'-5' polymerase implicated in noncoding RNA processing. (dissertation).
11. Eichinger, L., & Rivero, F. (2016). Dictyostelium discoideum protocols. Humana Press.
12. El Yacoubi B, Bailly M, de Crécy-Lagard V. Biosynthesis and function of posttranscriptional modifications of transfer RNAs. Annu Rev Genet. 2012;46:69-95. doi: 10.1146/annurev-genet-110711-155641. Epub 2012 Aug 16. PMID: 22905870.
13. Fey, P., Kowal, A. S., Gaudet, P., Pilcher, K. E., Chisholm, R. L. Protocols for growth and development of Dictyostelium discoideum. Nat Protoc. 2007; 2:1307-16
14. Fritze, C. E., & Anderson, T. R. (2000). Epitope tagging: General method for tracking recombinant proteins. Methods in Enzymology, 3–16. https://doi.org/10.1016/s0076-6879(00)27263-7
15. Giegé R, Sissler M, Florentz C 1998. Universal rules and idiosyncratic features in tRNA identity. Nucleic Acids Res 26: 5017–5035
16. Gerace E, Moazed D. Affinity Purification of Protein Complexes Using TAP Tags. Methods Enzymol. 2015;559:37-52. doi: 10.1016/bs.mie.2014.11.007. Epub 2015 Mar 29. PMID: 26096502; PMCID: PMC4794980.
17. Greenberg, J. R. (1979). Ultraviolet light-induced crosslinking of mrna to proteins. Nucleic Acids Research, 6(2), 715–732. https://doi.org/10.1093/nar/6.2.715
18. Gu W., Jackman J.E., Lohan A.J., Gray M.W., Phizicky E.M. tRNAHis maturation:An essential yeast protein catalyzes addition of a guanine nucleotide to the 5’ end of tRNAHis . Genes & Dev. 2003;17:2889–2901
19. Gu W, Hurto RL, Hopper AK, Grayhack EJ, Phizicky EM 2005. Depletion of Saccharomyces cerevisiae tRNA(His) guanylyltransferase Thg1p leads to uncharged tRNAHis with additional m(5)C. Mol Cell Biol25

20. Götze, M., Sarnowski, C. P., de Vries, T., Knörlein, A., Allain, F. H.-T., Hall, J., Aebersold, R., & Leitner, A. (2021). Single nucleotide resolution RNA–protein crosslinking mass spectrometry: A simple extension of the CLIR-MS Workflow. Analytical Chemistry, 93(44), 14626–14634. https://doi.org/10.1021/acs.analchem.1c02384

21. Hernan, R., Heuermann, K., & Brizzard, B. (2018). Multiple epitope tagging of expressed proteins for enhanced detection. BioTechniques, 28(4), 789–793. https://doi.org/10.2144/00284pf01

22. Himeno H, Hasegawa T, Ueda T, Watanabe K, Miura K-i, Shimizu M 1989. Role of the extra G-C pair at the end of the acceptor stem of $tRNA^{His}$ in aminoacylation. Nucleic Acids Res 17: 7855–7863

23. Hoffman EA, Frey BL, Smith LM & Auble DT Formaldehyde crosslinking: a tool for the study of chromatin complexes. J. Biol. Chem. 2015; 290, 26404–26411.

24. Houser-Scott, F., & Engelke, D. R. (2001). Protein–RNA interactions. Encyclopedia of Life Sciences. https://doi.org/10.1038/npg.els.0001360

25. Howell, N. W., & Jackman, J. E. (2019). Impact of chemical modification on function. Encyclopedia of Life Sciences, 1–11. https://doi.org/10.1002/9780470015902.a0028527

26. Hyde SJ, Eckenroth BE, Smith BA, Eberley WA, Heintz NH, Jackman JE, Doublie S 2010. tRNAHis guanylyltransferase (THG1), a unique 3'-5' nucleotidyl transferase, shares unexpected structural homology with canonical 5'-3' DNA polymerases. Proc Natl Acad Sci 107: 20305–20310

27. Immunoprecipitation of flag fusion proteins using monoclonal antibody affinity gels. (n.d.). https://www.sigmaaldrich.com/US/en/technical-documents/protocol/protein-biology/protein-purification/flag-immuno-precipitation-protocol#peptide

28. Jackman, J. E. (2010). tRNA biogenesis. Encyclopedia of Life Sciences. https://doi.org/10.1002/9780470015902.a0020894

29. Jackman JE, Gott JM, Gray MW. Doing it in reverse: 3'-to-5' polymerization by the Thg1 superfamily. RNA. 2012 May;18(5):886-99. doi: 10.1261/rna.032300.112. Epub 2012 Mar 28. PMID: 22456265; PMCID: PMC3334698.

30. Jackman JE, Phizicky EM 2006b. tRNAHis guanylyltransferase catalyzes a 3'-5' polymerization reaction that is distinct from G_{-1} addition. Proc Natl Acad Sci 103: 8640–8645

31. Jackson, V. (1978). Studies on histone organization in the nucleosome using formaldehyde as a reversible crosslinking agent. Cell, 15(3), 945–954. https://doi.org/10.1016/0092-8674(78)90278-7

32. Jahn D., Pande S. Histidine tRNA guanylyltransferase from Saccharomyces cerevisiae. II. Catalytic mechanism. J. Biol. Chem. 1991;266:22832–22836.

33. Kennedy-Darling, J., & Smith, L. M. (2014). Measuring the formaldehyde protein–DNA crosslink reversal rate. Analytical Chemistry, 86(12), 5678–5681. https://doi.org/10.1021/ac501354y

34. Lonergan KM, Gray MW 1993a. Editing of transfer RNAs in Acanthamoeba castellanii mitochondria. Science 259: 812–816

35. Lonergan KM, Gray MW 1993b. Predicted editing of additional transfer RNAs in Acanthamoeba castellanii mitochondria. Nucleic Acids Res21: 4402 doi: 10.1093/nar/21.18.4402

36. Long, Y., Abad, M. G., Olson, E. D., Carrillo, E. Y., & Jackman, J. E.. Identification of distinct biological functions for four 3’-5’ RNA polymerases. Nucleic Acids Research. 2016; 44(17), 8395–8406. https://doi.org/10.1093/nar/gkw681

37. Marck C., Grosjean H. tRNomics: Analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. RNA. 2002;8:1189–1232

38. Meisenheimer KM, Koch TH. Photocrosslinking of nucleic acids to associated proteins. Crit Rev Biochem Mol Biol 1997, 32: 101–140.

39. Nameki N, Asahara H, Shimizu M, Okada N, Himeno H 1995. Identity elements of Saccharomyces cerevisiae $tRNA^{His}$. Nucleic Acids Res 23: 389–394

40. Ogawa S, Yoshino R, Angata K, Iwamoto M, Pi M, Kuroe K, Matsuo K, Morio T, Urushihara H, Yanagisawa K, et al. 2000. The mitochondrial DNA of Dictyostelium discoideum: complete sequence, gene content and genome organization. Mol Gen Genet 263: 514–519

41. Price D.H., Gray M.W. A novel nucleotide incorporation activity implicated in the editing of mitochondrial transfer RNAs in Acanthamoeba castellanii. RNA. 1999;5:302–317.

42. Ramanathan, M., Porter, D. F., & Khavari, P. A. (2019). Methods to study RNA–protein interactions. Nature Methods, 16(3), 225–234. https://doi.org/10.1038/s41592-019-0330-1

43. Rao BS, Maris EL, Jackman JE 2010. tRNA 5'-end repair activities of tRNAHis guanylyltransferase (Thg1)-like proteins from Bacteria and Archaea. Nucl Acids Res doi: 10.1093/nar/gkq976

44. Rosen AE, Musier-Forsyth K 2003. Recognition of G-1:C73 atomic groups by Escherichia coli histidyl-tRNA synthetase. J Am Chem Soc126: 64–65

45. Smith BA, Jackman JE. Kinetic Analysis of 3'-5' Nucleotide Addition Catalyzed by Eukaryotic tRNA(His) Guanylyltransferase. Biochemistry. 2012 Jan 10;51(1):453-65.

46. Sprinzl M., Horn C., Brown M., Ioudovitch A., Steinberg S. Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. 1998;26:148–153.

47. Ule, J., Jensen, K., Mele, A., & Darnell, R. B. (2005a). Clip: A method for identifying protein–RNA interaction sites in living cells. Methods, 37(4), 376–386. https://doi.org/10.1016/j.ymeth.2005.07.018

48. Valášek, L., Szamecz, B., Hinnebusch, A. G., & Nielsen, K. H. (2007). In vivo stabilization of preinitiation complexes by formaldehyde crosslinking. Methods in Enzymology, 163–183. https://doi.org/10.1016/s0076-6879(07)29008-1

49. Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B, Blue SM, Nguyen TB, Surka C, Elkins K, Stanton R, Rigo F, Guttman M, Yeo GW. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat Methods. 2016 Jun;13(6):508-14. doi: 10.1038/nmeth.3810. Epub 2016 Mar 28. PMID: 27018577; PMCID: PMC4887338.

50. Waldron, C., & Lacroute, F. (1975). Effect of growth rate on the amounts of ribosomal and transfer ribonucleic acids in yeast. Journal of Bacteriology, 122(3), 855–865. https://doi.org/10.1128/jb.122.3.855-865.1975

51. Wheeler EC, Van Nostrand EL, Yeo GW. Advances and challenges in the detection of transcriptome-wide protein-RNA interactions. Wiley Interdiscip Rev RNA. 2018 Jan;9(1):e1436. doi: 10.1002/wrna.1436. Epub 2017 Aug 29. PMID: 28853213; PMCID: PMC5739989.