Substrate Recognition and Specificity of a Minimal Protein-Only RNase P

by

Catherine A. Wilhelm

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
at the University of Michigan
2024

Doctoral Committee:

      Assistant Professor Markos Koutmos, Chair
      Associate Professor Kristin Koutmou
      Professor Neil Marsh
      Associate Professor Randy Stockbridge

Catherine A. Wilhelm

cawilhel@umich.edu

ORCID iD: 0000-0002-4163-9658

Dedication

This dissertation is dedicated to my family, friends, and my past, present, and future self.

Acknowledgements

Graduate school is one of the hardest things I've done, further complicated by the external challenges I faced throughout. I would not have been able to finish my degree without the help of the following people:

Table of Contents

List of Tables

List of Figures

Abstract

tRNAs are typically transcribed with extended 5' and 3' ends that must be removed before they attain their active form. One of the first steps of tRNA processing in nearly every organism is the removal of the 5' leader sequence by ribonuclease P (RNase P). This work investigates a newly discovered class of RNase P enzymes, Homologs of *Aquifex* RNase P (HARPs). In contrast to other RNase Ps, HARPs consist only of a metallonuclease domain and lack the substrate recognition domains essential in other classes of RNase P enzymes. Determination of the cryo-EM structure of *A. aeolicus* HARP (Aq880) and two crystal structures of *H. thermophilus* HARP (Hth1307) reveal that both enzymes form large ring-like assemblies: a dodecamer in Aq880 and a tetradecamer in Hth1307. In both oligomers, the enzyme active site is 42 Å away from a positively charged helical region, as seen in other protein-only RNase P enzymes, which likely serves to recognize and bind the elbow region of the pre-tRNA substrate. In addition, native mass spectrometry (MS) was used to confirm and characterize the previously unreported tetradecamer state. Notably, multiple oligomeric states of Hth1307 are able to cleave pre-tRNAs. Furthermore, single-turnover kinetic studies indicate that Hth1307 cleaves pre-tRNAs from multiple species with a preference for native substrates. These data provide a closer look at the nuanced similarities and differences in tRNA processing across disparate classes of RNase P.

Chapter 1 Introduction

The Central Dogma of Biology states that DNA is transcribed into RNA which is translated into proteins. This is a gross oversimplification of the dynamic processes within cells that regulate gene expression and overall cellular function, neglecting the processes of post-transcriptional and post-translational modification and maturation, as well as the contributions of non-coding RNAs (**Figure 1.1***A*). One such non-coding RNA is transfer RNA (tRNA), which is used to decode messenger RNA during translation to form chains of amino acids. The process of producing a fully mature tRNA that can be used in translation is lengthy, requiring cleavage of excess nucleotides, the addition of a 3' CCA tail to tRNAs that do not natively encode one, modification of nucleotides at specific positions, and aminoacylation at the 3' end (**Figure 1.1***B*).

1.1 Processing of the 5' end of precursor tRNAs by Ribonuclease P

Transfer RNAs are transcribed *in vivo* as precursor tRNAs (pre-tRNAs) with extra sequences at their 5' and 3' termini. These additional sequences must be cleaved to produce mature tRNA molecules that are further modified and used in translation of mRNA into protein. The 5' leader sequence removal is catalyzed by Ribonuclease P (RNase P) in all organisms (except *N. equitans*, which does not include a 5' leader in its tRNA genes) (1). The discovery of RNA-dependent RNase P by Altman and colleagues was one of the first discovered ribozymes and led to a share of the Nobel Prize for RNA-dependent catalysts (2–5). The first indications of an RNA-independent protein-only RNase P (PRORP) came from work in 1998, which demonstrated human mitochondrial RNase P (*Hs* (mt)RNase P) was resistant to nuclease activity (6). However, the molecular constituents responsible for activity remained undefined until a decade later, when Holzmann et al. (2008) identified and reconstituted the components of *Hs* (mt)RNase P through an extensive combination of proteomic and biochemical techniques (7). Following the discovery of metazoan mitochondrial PRORP, bioinformatic searches identified three PRORP paralogs in *Arabidopsis thaliana* (*At*). *At* PRORP1 localizes to the mitochondria and chloroplasts, while *At* PRORP2 and *At* PRORP3 are localized to the nucleus. Most recently, a novel form of PRORP was discovered in bacteria and archaea, termed homologs of *Aquifex* RNase P (HARPs) (8).
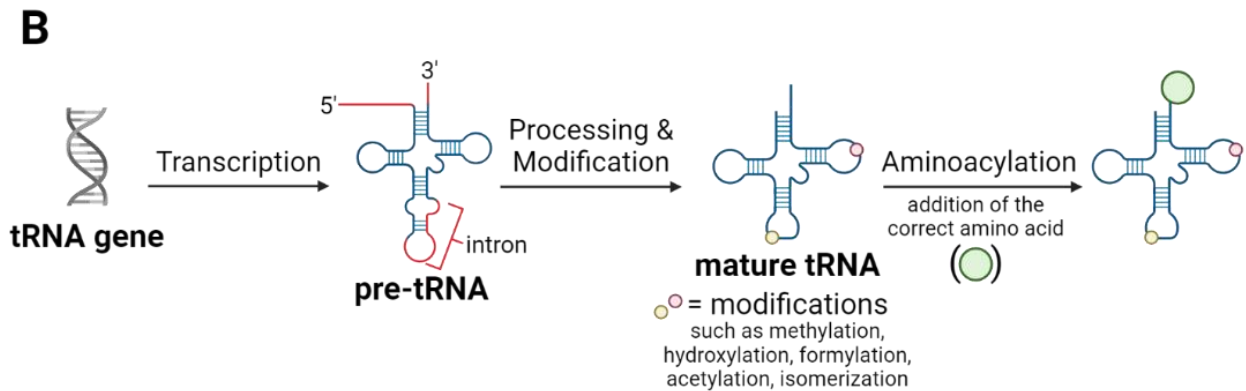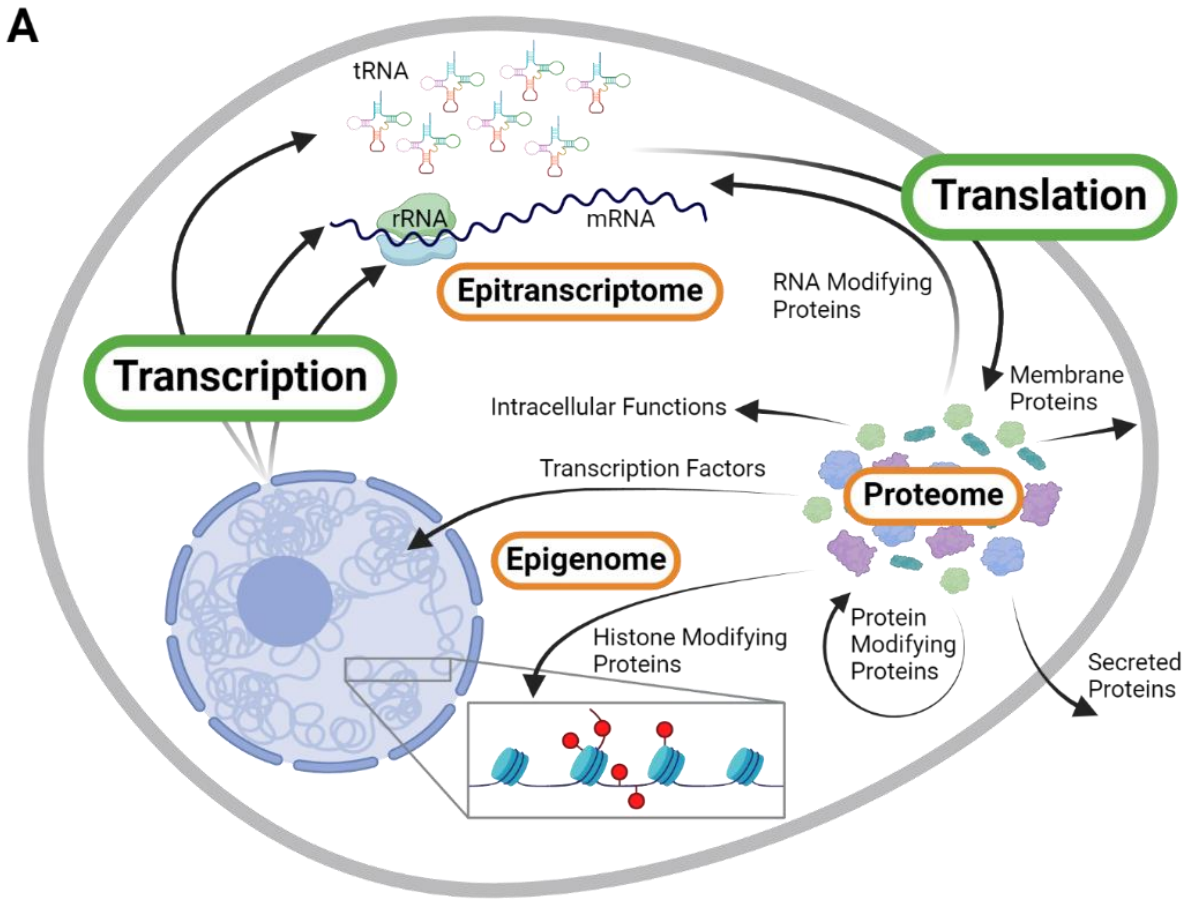
Figure 1.1: The Central Dogma of Biology and a summary of tRNA processing.

A. The "circular" central dogma of biology, which expands on the variety of processing and modification events that take place compared to the traditional central dogma of biology. **B.** The maturation process of a transfer RNA. Precursor tRNA is transcribed from a tRNA gene with additional nucleotide sequences at the 5' and 3' ends, sometimes also containing an intron. After cleavage and splicing, tRNAs are heavily modified. Once fully matured, the tRNA is aminoacylated by the aminoacyl-tRNA synthetase corresponding to the anticodon of the tRNA.

1.1.1 Biological Relevance of Ribonuclease P

1.1.1.1 Pre-tRNA processing by human mitochondrial RNase P is essential for human health

Mitochondrial tRNA ((mt)tRNA) processing is essential for cellular health. Mitochondria encode the mRNA, rRNA, and tRNA genes required to synthesize the proteins of the electron transport chain and produce energy in the cell. These genes are transcribed as three polycistronic transcripts in which tRNA genes are distributed among rRNA and mRNA genes, often without any noncoding sequence present between the tRNA and coding sequences (9). Therefore, defective (mt)tRNA maturation can impair the processing of other types of RNA molecules and negatively impact mitochondrial function (10, 11). Defects in (mt)tRNA maturation have been shown to reduce steady-state levels of mitochondrial ATP and mature (mt)tRNAs, rates of mitochondrial polypeptide synthesis, activity of cellular respiration chains, and membrane potentials across the mitochondrial membrane. Moreover, erroneous aminoacylation events have been observed in cytoplasmic hybrid (cybrid) cells defective in 5' pre-tRNA cleavage (12, 13).

Patient mutations in the genes encoding the proteins of *Hs* (mt)RNase P complex and in the mitochondrial DNA sequences encoding its substrate tRNAs have been linked to numerous mitochondrial diseases, several of which have no known cure and can lead to premature mortality (14) (**Figure 1.2**). Such diseases include maternally inherited essential hypertension, MELAS (mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes), 2-methyl-3-hydroxybutyric aciduria (HSD10), and ophthalmoplegia (12, 15–25). Several studies have been conducted on patient mutations to date. For example, fibroblasts derived from two patients presenting with lactic acidosis contained missense mutations in the TRMT10C gene (25). These mutant fibroblasts exhibited decreased levels of TRMT10C, as well as an accumulation of (mt)pre-tRNA, indicative of impaired tRNA maturation and protein synthesis in the mitochondria (25).
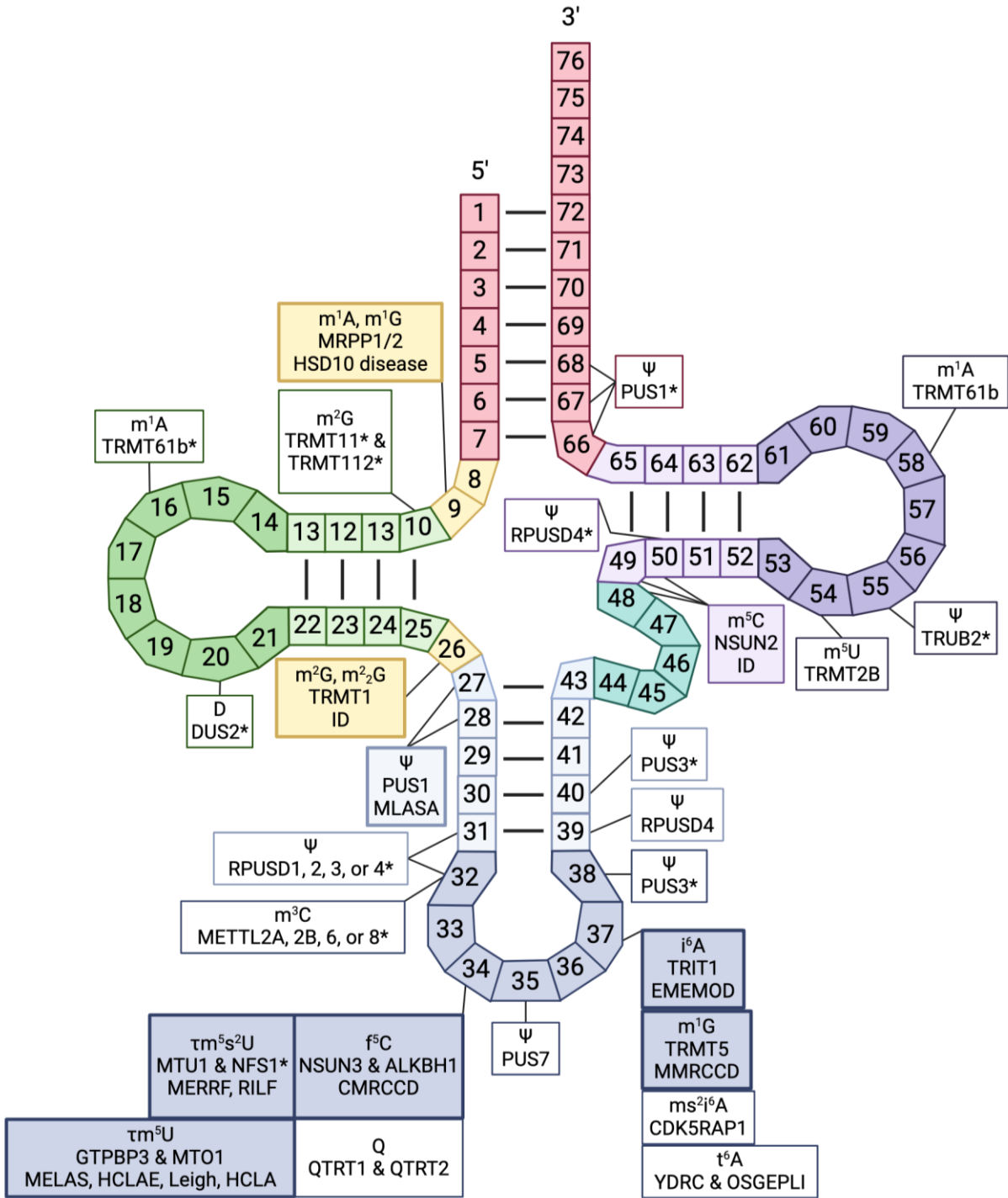
Figure 1.2: Map of mitochondrial tRNA mutations and diseases.

A cartoon representation of the secondary structure of mitochondrial tRNA highlighting modifications, the enzymes that install them, and diseases associated with tRNA or modifying enzyme mutations. Modifications with no known disease mutations are in white boxes, while modifications with diseases linked to tRNA or modifying enzyme mutants are in colored boxes. Enzymes with an * are hypothesized to install the mutation but have not been proven. Data was compiled from (26).

Notably, these mitochondrial defects were rescued by lentiviral transduction of the wild-type TRMT10C gene into the patient-derived fibroblasts, highlighting the role that the mutant form of the gene plays in pathogenesis (25).

Another study focused on the impact of 11 disease-linked mutations in three (mt)pre-tRNAs on *Hs* (mt)RNase P activity and binding affinity (27). These mutations showed varying effects on *Hs* (mt)RNase P binding and activity, with some mutants affecting binding only, cleavage only, methylation only, or a combination of the three (27). Such studies point to a linkage between *Hs* (mt)RNase P and various mitochondrial diseases; however, much remains to be elucidated about the mechanisms of pathogenesis underlying these illnesses, which would provide the basis for future studies into how these diseases and their associated symptoms may be treated.

1.1.1.2 Use of Arabidopsis thaliana protein-only Ribonuclease P against plant viruses

A. thaliana PRORPs are promiscuous, cleaving the 5' end of mitochondrial, chloroplast, and nuclear pre-tRNAs (28). There is also evidence that *At* PRORPs play a role in mRNA maturation by cleaving tRNA-like structures (TLS) at the 5' or 3' ends in some plant mitochondria, as demonstrated in both in vitro and in vivo systems (29). Some positive-strand RNA plant viral genomes contain 3'-tRNA-like structures, such as Turnip Yellow Mosaic Virus (TYMV) and Oilseed Rape Mosaic Virus (ORMV) (29, 30). These TLS are crucial in viral infectivity and viral replication, making them a potential target for treatments to provide viral resistance to plants and protect food sources. Taking advantage of the promiscuity of *At* PRORPs, a ribonuclease termed CytoRP was engineered to localize to the cytosol and protect against plant viruses by cleaving TLS in viral genomes such as TYMV and ORMV. This technology holds great promise as an extensive defense against multiple viruses containing TLS, thus preventing crop yield loss (29).

1.1.2 Structural characterization of PRORP from human mitochondria and A. thaliana

Protein-only RNase P is prevalent throughout eukaryotic species. Although many PRORPs have been putatively identified through sequence homology, biochemical and structural characterization of eukaryotic PRORPs is limited. Currently, our understanding of eukaryotic PRORP is based on data from only two species, *A. thaliana* and *H. sapiens*. *At* PRORP isoforms have been used as a model for understanding *Hs* (mt)RNase P because they have high homology and the *Hs* (mt)RNase P complex is difficult to purify.

1.1.2.1 Molecular architecture of A. thaliana PRORP

The first structures were obtained via x-ray crystallography using *At* PRORP1 and *At* PRORP2. These structures revealed that eukaryotic PRORPs form a V-like structure that consists of three domains: an N-terminal pentatricopeptide repeat (PPR) domain, a central linker domain, and a C-terminal metallonuclease domain (**Figure 1.3*A***) (31). The PPR domain is made up of approximately five PPR motifs, each consisting of about 35 amino acids that contain RNA-interacting residues and acts as the tRNA recognition domain (7, 32). The central linker domain harbors a zinc-binding site that is highly conserved across PRORPs. Zinc coordination both stabilizes the central domain and promotes the proper orientation of the PPR domain with respect to the metallonuclease domain that allows binding of pre-tRNA, as supported by crystal structures(33–35). The central linker domain also contains a positively charged region near the zinc binding site, possibly allowing for interactions with pre-tRNA (36). Lastly, the NYN (Nedd4-BP1, YacP nuclease)/PIN-like metallonuclease domain carries out the 5' endonucleolytic cleavage of pre-tRNAs. Conserved acidic residues, specifically aspartates, in the active site coordinate divalent magnesium ions necessary for catalysis (**Figure 1.3*B***) (37).
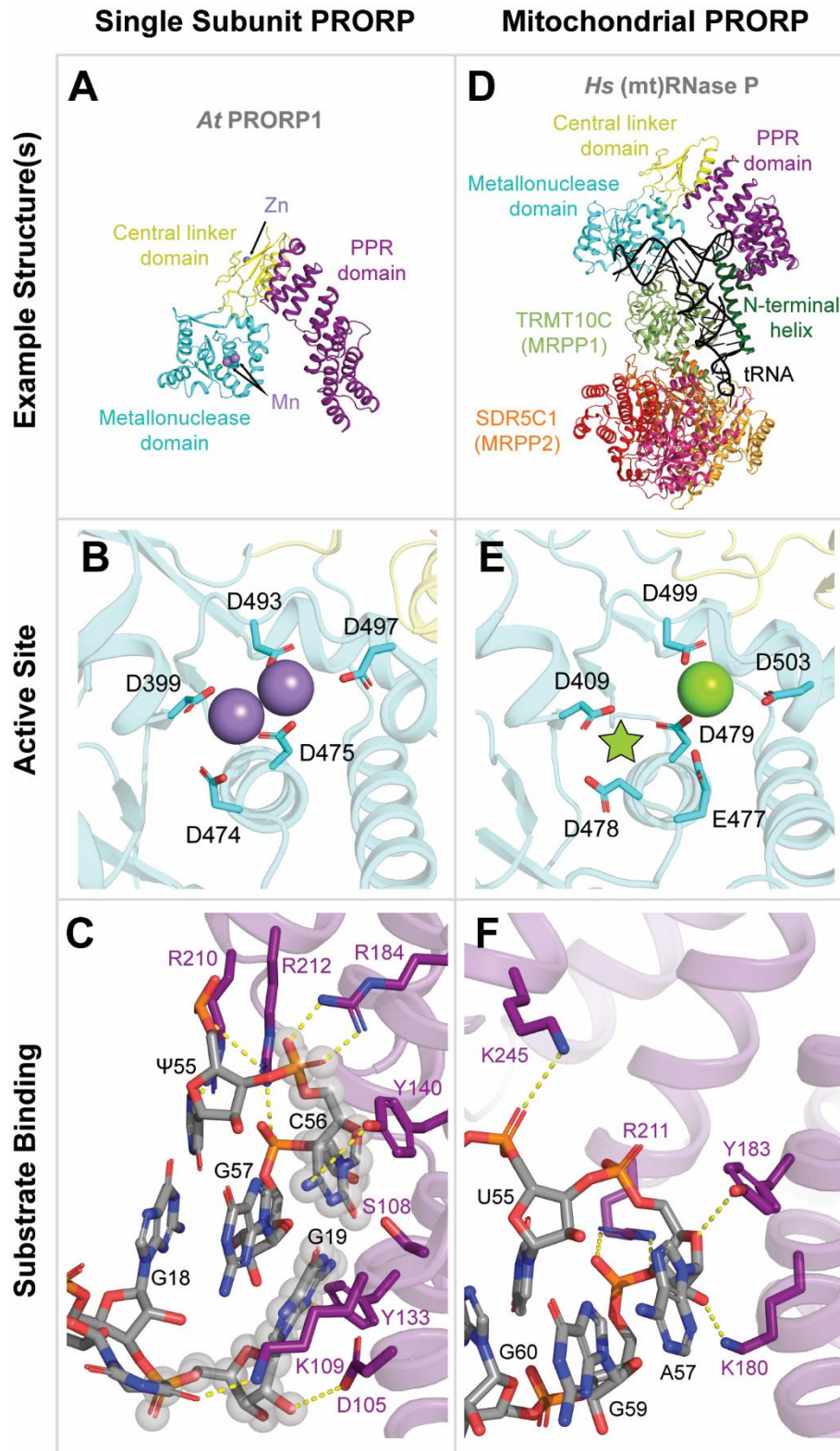
Figure 1.3: Comparison of *A. thaliana* and *H. sapiens* PRORP structures.

A. Crystal structure of *At* PRORP1 (PDBID: 4G24). The metallonuclease, central linker, and PPR domains are shown in cyan, yellow, and dark purple, respectively. The two manganese ions in the active site of the metallonuclease

domain correspond to the two metal ions used in catalysis. **B.** Zoomed in active site of *At* PRORP1 (PDBID: 4G24) including two manganese ions. **C.** Yeast tRNA[Phe] and the PPR domain of *At* PRORP1 (PDBID: 6LVR). The conserved G19-C56 base pair is highlighted with transparent grey spheres. Note the base stacking interaction between the C56 base and the phenol ring of Y140. **D.** Cryo-EM structure of *Hs* (mt)RNase P (PDBID: 7ONU), with the *Hs* PRORP subunit sharing the same coloring as *At* PRORP1 in A. The TRMT10C subunit is shown in shades of green. The four SDR5C1 subunits are shown in shades of red, pink, and orange. The pre-tRNA substrate is *Hs* (mt)tRNA[Tyr] with a two-nucleotide leader sequence. **E.** Zoomed in active site of *Hs* (mt)RNase P (PDBID: 7ONU) including one calcium ion, with a green star representing the proposed binding site of an additional metal ion. **F.** Human (mt)pre-tRNA[Tyr] and the PPR domain of *Hs* (mt)RNase P (PDBID: 7ONU). The conserved Y183 residue (Y140 in *At* PRORP1) forms a hydrogen bond with the substrate rather than the stacking interaction seen in **C**.

The X-ray structure of the *At* PRORP1 PPR domain bound to yeast tRNA[Phe] revealed interactions important for substrate recognition, which are supported by additional kinetic and binding assays conducted with mutant *At* PRORP1 (35). The tRNA-bound PPR domain forms a curved structure with an inner concave surface that accommodates the 'elbow' region formed by the D and TψC loops of the tRNA. Key interactions are formed between residues in the PPR motif and conserved nucleotides and structural elements of the tRNA (**Figure 1.3***C*). Notably, a conserved G19–C56 tertiary interaction between the D and TψC loops serves as a key element of substrate recognition, forming multiple interactions with PPR domain residues. Recognition of this G19-C56 base pair is a common strategy for tRNA recognition and is shared with RNA-based RNase P (38, 39).

1.1.2.2 Molecular architecture of H. sapiens mitochondrial PRORP

In non-metazoan eukaryotes, PRORP demonstrates efficient processive activity as a single subunit. In contrast, an orthologue of PRORP found in the mitochondria of metazoan eukaryotes has been shown to exist as a member of the tripartite *Hs* (mt)RNase P (**Figure 1.3***D*). This complex consists of the PRORP orthologue, often referred to as mitochondrial RNase P protein 3 (*Hs* PRORP, aka MRPP3), and two accessory proteins, known as TRMT10C (MRPP1) and SDR5C1 (MRPP2) (7). MRPP3 behaves as the catalytic center of the complex to perform 5' endonucleolytic cleavage of its substrate pre-tRNAs with an active site nearly identical to that of *At* PRORP (**Figure 1.3***E*) (36). TRMT10C is a methyltransferase that requires SDR5C1, a dehydrogenase, for

methylation activity. The TRMT10C-SDR5C1 subcomplex consists of a single TRMT10C molecule and four SDR5C1 molecules that interact via protein-protein interactions (40, 41). Together, the TRMT10C-SDR5C1 subcomplex can methylate purine nucleotides found at the 9th position in nineteen out of twenty-two *Hs* (mt)tRNAs, without requiring the presence of *Hs* PRORP *in vitro* (40, 42). It is hypothesized that the subcomplex serves as a scaffold for *Hs* PRORP, stabilizing the active site conformation of the endonuclease (7, 41, 43).

The recently published cryo-EM structure of *Hs* (mt)RNase P in complex with (mt)pre-tRNA^{Tyr} (**Figure 1.3D**) provides evidence in support of this hypothesis. Upon binding, the pre-tRNA substrate is positioned within the cleft of the "V" conformation, where the tRNA elbow region interacts with the PPR domain while the scissile phosphodiester bond of the pre-tRNA is positioned within the active site of the metallonuclease domain. The distance between the binding and catalytic sites in RNase P enzymes is 42 Å and highly conserved in both ribozyme and protein-only RNase P enzymes. This distance is a "molecular ruler" to accommodate the substrate tRNA (44).

One of the striking features of this structure is the extensive contact of the TRMT10C-SDR5C1 subcomplex with the (mt)pre-tRNA body. The interactions between the substrate and protein differ between *At* PRORP and *Hs* (mt)RNase P, with the latter lacking the base stacking interactions seen with *At* PRORP (**Figure 1.3F**). The TRMT10C–SDR5C1 subcomplex interacts with all four arms of the pre-tRNA, forming a shape-complementary scaffold around the tertiary L-shape tRNA structure. The C-terminal methyltransferase domain of TRMT10C interacts with the anticodon and D stem of the pre-tRNA, positioning the m^{1}R9 methyl acceptor within its active site. Additionally, the acceptor stem of the pre-tRNA interacts with the top of the methyltransferase domain, where it is posited to recruit *Hs* PRORP to the TRMT10C-SDR5C1 subcomplex for 5'

cleavage (40). The methyltransferase domain is linked to the N-terminal domain (NTD) of TRMT10C through an 'adapter helix' which interacts with the SDR5C1 tetramer and bases C32, U33, and U35 of the anticodon loop. The NTD of TRMT10C is comprised of an alpha connector helix which is positioned opposite the methyltransferase domain between the D arm and anticodon arm of the pre-tRNA. The interactions of this helix appear to result in altered conformations of the tRNA body, exhibiting slight displacements of the D arm and anticodon loop in comparison to canonical tRNA structures. In addition to interactions with the tRNA, the NTD connector helix also interacts with the PPR domain of *Hs* PRORP, aiding its recognition of the tRNA elbow region. This aid in recognition of the tRNA elbow region could be of particular importance, as the D-loop of the tRNA is unresolved in this structure, suggesting an absence of the conserved structural interactions that canonically define this region. Taken together, the extensive interactions between the TRMT10C-SDR5C1 subcomplex and the tRNA body, along with the requirement of TRMT10C-SDR5C1 subcomplex for efficacious 5' cleavage, support a proposed scaffolding role for TRMT10C and SDR5C1.
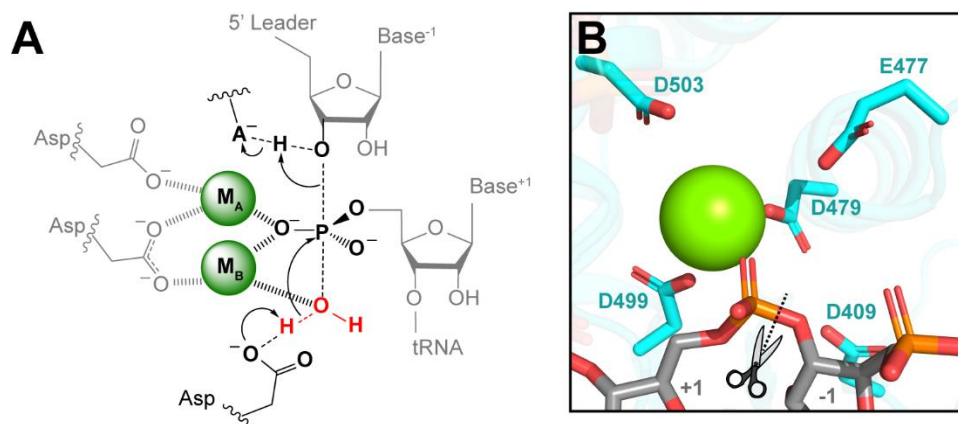
Figure 1.4: Mechanism of PRORP.

A. Conserved mechanism of PRORPs. In this mechanism, one Mg^{2+}(B) activates a coordinated water for nucleophilic attack of the scissile phosphodiester bond, as the second Mg^{2+}(A) is proposed to stabilize the developing charge on the 3'-oxyanion leaving group. **B.** Zoomed in active site of *Hs* (mt)RNase P with (mt)pre-tRNA[Tyr] substrate (PDBID: 7ONU) to show scissile bond between the +1 and -1 nucleotides denoted by the scissors and dotted line.

11

1.1.3 The conserved mechanism of protein-only Ribonuclease P

All PRORP enzymes contain a metallonuclease domain belonging to the PilT N-terminal (PIN) domain-like superfamily that catalyzes phosphodiester bond cleavage. This domain forms an exposed active site cleft containing two metal ions that are coordinated by a conserved set of 4 aspartate residues. Due to the shared metallonuclease domain, PRORPs are proposed to utilize the same chemical mechanism. Most of our mechanistic understanding of PRORPs comes from work done with the eukaryotic *At* PRORP paralogs. Metal ion and pH dependency studies of *At* PRORP support phosphodiester bond hydrolysis catalyzed by a classical two-metal ion mechanism (**Figure 1.4A-B**) (45). This mechanism is presumed to be conserved in *Hs* (mt)RNase P based on the conservation of the structure of the active site when compared to *At* PRORPs (**Figure 1.3B, E**).

1.1.4 Functional characterization of canonical protein-only ribonuclease P

1.1.4.1 Eukaryotic PRORPs demonstrate substrate specificity

All three *At* PRORPs were found to have less than four-fold difference in binding affinity for *At* (mt)pre-tRNA^{Cys}, *At* (chlor)pre-tRNA^{Phe}, *At* (nuc)pre-tRNA^{Cys}, *At* (nuc)pre-tRNA^{Phe}, suggesting they use similar binding modes for pre-tRNA substrate recognition (46). However, and somewhat contradictorily, *At* PRORP2 was found to have higher binding affinity for *Synechocystis* pre-tRNA^{Gln} than *At* PRORP3, although the binding affinities for *At* (chlor)pre-tRNA^{Arg} was similar (47). This suggests some difference in *At* PRORP2 and *At* PRORP3 substrate scope or recognition. In addition, *At* PRORP2 has low activity at high temperatures while *At* PRORP3 remains active. Based on this information, it was proposed that *At* PRORP2 and *At* PRORP3 may have functional synergy in the nucleus, with *At* PRORP3 acting as a "safety net" for tRNA maturation under stress conditions (47).

12

Work has shown *At* PRORP1 and *At* PRORP2 do not bind and process most human mitochondrial pre-tRNAs efficiently (42). Thus, it is proposed *Hs* (mt)RNase P uses different strategies or interactions for recognizing its substrates compared to the single component *At* PRORP1 and 2 enzymes (42). *Hs* (mt)RNase P has higher activity and binding affinities for (mt)pre-tRNA^{Ile} and (mt)pre-tRNA^{Leu} compared to other native substrates, highlighting some sort of substrate preference or specificity (42, 48). However, much like *At* PRORPs, the cause of such specificity is unknown.

1.1.4.2 Human mitochondrial PRORP functions independently of TRMT10C and SDR5C1

It was long assumed that *Hs* (mt)RNase P required the TRMT10C/SDR5C1 subcomplex for activity. Recent work has demonstrated that *Hs* PRORP can independently cleave some mitochondrial pre-tRNA substrates (pre-tRNA^{Ala}, pre-tRNA^{His}, pre-tRNA^{Met}) and *T. thermophilus* pre-tRNA^{Gly}, although a small portion of pre-tRNA^{Gly} was miscleaved at positions -1 and -2 in the absence of the TRMT10C/SDR5C1 subcomplex. It is worth noting that *Hs* PRORP alone was not capable of cleaving all mitochondrial pre-tRNA substrates, and there is no clear correlation between structural features of the pre-tRNAs and the requirement of the TRMT10C/SDR5C1 subcomplex for cleavage. The binding affinity of *Hs* PRORP alone was nearly equivalent to the holo complex, but the observed rate decreased 10-fold. It is proposed that the role of the TRMT10C/SDR5C1 subcomplex is to aid in aligning the pre-tRNA substrate for correct cleavage at the +1 position by mediating a conformational change, which is the rate-limiting step in the reaction (48). The model described refers to a conformational change in the Hs PRORP protein, but it does not account for the possibility of a conformational change to the pre-tRNA substrate mediated by the TRMT10C/SDR5C1 subcomplex. More work is needed to understand why some

13

(mt)pre-tRNA substrates require the TRMT10C/SDR5C1 subcomplex for cleavage by *Hs* PRORP and what exact conformational change is facilitated by the TRMT10C/SDR5C1 subcomplex (45).

1.2 Discovery of a novel protein-only ribonuclease P

Aquifex aeolicus remained a mystery in the field of ribonuclease P for decades. When the *A. aeolicus* genome was first published in 2001, no identifiable bacterial ribonuclease P gene was found (49). However, tRNA gene clustering in the genome, the presence of 5'-end-matured tRNA in lysates, and the presence of RNase P-like activity in lysates indicated tRNAs in *A. aeolicus* must undergo some kind of 5' end processing (50, 51). It was not until 2017 that the protein responsible for this activity, Aq880, was identified. Bioinformatic analysis revealed homologous proteins in several archaea and bacteria, so the class of enzymes was termed "Homologs of *Aquifex* RNase P" (HARP) (8). Aq880 was determined to be significantly smaller than canonical PRORPs, at only ~200 amino acids and ~23 kDa compared to ~600 amino acids and ~65 kDa respectively. Sequence alignments of all PRORPs indicate the active site of HARPs is similar to the conserved active sites in *At* and *Hs* PRORPs, suggesting HARPs share the same catalytic mechanism as previously characterized PRORPs. However, HARPs seem to lack the canonical pentatricopeptide repeat region that serves as the substrate recognition domain in other PRORPs.

1.3 Research Goals

The following work seeks to address outstanding questions regarding the novel class of protein-only RNase P enzymes called HARPs. **Chapter 1** has laid the groundwork for our current understanding of PRORPs from *A. thaliana* and *H. sapiens* mitochondria. **Chapter 2** details the solution of structures of HARP from *A. aeolicus* and *H. thermophilus* and subsequent analysis and

comparisons. **Chapter 3** details the use of kinetic assays to understand how the activity of HARPs

aligns and differs from previously characterized PRORPs.

1.4 References

1. Randau, L., Schröder, I., and Söll, D. (2008) Life without RNase P. *Nature*. **453**, 120–123

2. Altman, S. (2000) The road to RNase P. *Nat Struct Biol*. **7**, 827–828

3. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*. **35**, 849–857

4. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell*. **31**, 147–157

5. Pinker, F., Bonnard, G., Gobert, A., Gutmann, B., Hammani, K., Sauter, C., Gegenheimer, P. A., and Giegé, P. (2013) PPR proteins shed a new light on RNase P biology. *RNA Biol*. **10**, 1457–1468

6. Rossmanith, W., and Karwan, R. M. (1998) Characterization of Human Mitochondrial RNase P: Novel Aspects in tRNA Processing. *Biochemical and Biophysical Research Communications*. **247**, 234–241

7. Holzmann, J., Frank, P., Löffler, E., Bennett, K. L., Gerner, C., and Rossmanith, W. (2008) RNase P without RNA: Identification and Functional Reconstitution of the Human Mitochondrial tRNA Processing Enzyme. *Cell*. **135**, 462–474

8. Nickel, A. I., Wäber, N. B., Gößringer, M., Lechner, M., Linne, U., Toth, U., Rossmanith, W., and Hartmann, R. K. (2017) Minimal and RNA-free RNase P in Aquifex aeolicus. *Proceedings of the National Academy of Sciences*. **114**, 11121–11126

9. Ojala, D., Montoya, J., and Attardi, G. (1981) tRNA punctuation model of RNA processing in human mitochondria. *Nature*. **290**, 470–474

10. Levinger, L., Oestreich, I., Florentz, C., and Mörl, M. (2004) A pathogenesis-associated mutation in human mitochondrial tRNA[Leu(UUR)] leads to reduced 3'-end processing and CCA addition. *J Mol Biol*. **337**, 535–544

11. Wittenhagen, L. M., and Kelley, S. O. (2003) Impact of disease-related mitochondrial mutations on tRNA structure and function. *Trends in Biochemical Sciences*. **28**, 605–611

12. Jiang, P., Wang, M., Xue, L., Xiao, Y., Yu, J., Wang, H., Yao, J., Liu, H., Peng, Y., Liu, H., Li, H., Chen, Y., and Guan, M.-X. (2016) A Hypertension-Associated tRNA[Ala] Mutation Alters tRNA Metabolism and Mitochondrial Function. *Mol Cell Biol*. **36**, 1920–1930

13. Li, R., Liu, Y., Li, Z., Yang, L., Wang, S., and Guan, M.-X. (2009) Failures in mitochondrial tRNAMet and tRNA[Gln] metabolism caused by the novel 4401A>G mutation are involved in essential hypertension in a Han Chinese Family. *Hypertension*. **54**, 329–337

14. Hochberg, I., Demain, L. A. M., Richer, J., Thompson, K., Urquhart, J. E., Rea, A., Pagarkar, W., Rodríguez-Palmero, A., Schlüter, A., Verdura, E., Pujol, A., Quijada-Fraile, P., Amberger, A., Deutschmann, A. J., Demetz, S., Gillespie, M., Belyantseva, I. A., McMillan, H. J., Barzik, M., Beaman, G. M., Motha, R., Ng, K. Y., O'Sullivan, J., Williams, S. G., Bhaskar, S. S., Lawrence, I. R., Jenkinson, E. M., Zambonin, J. L., Blumenfeld, Z., Yalonetsky, S., Oerum, S., Rossmanith, W., Genomics England Research Consortium, Yue, W. W., Zschocke, J., Munro, K. J., Battersby, B. J., Friedman, T. B., Taylor, R. W., O'Keefe, R. T., and Newman, W. G. (2021) Bi-allelic variants in the mitochondrial RNase P subunit PRORP cause mitochondrial tRNA processing defects and pleiotropic multisystem presentations. *Am J Hum Genet*. **108**, 2195–2204

15. Bindoff, L. A., Howell, N., Poulton, J., McCullough, D. A., Morten, K. J., Lightowlers, R. N., Turnbull, D. M., and Weber, K. (1993) Abnormal RNA processing associated with a novel tRNA mutation in mitochondrial DNA. A potential disease mechanism. *J Biol Chem*. **268**, 19559–19564

16. Brandon, M. C., Lott, M. T., Nguyen, K. C., Spolim, S., Navathe, S. B., Baldi, P., and Wallace, D. C. (2005) MITOMAP: a human mitochondrial genome database—2004 update. *Nucleic Acids Res*. **33**, D611–D613

17. Chatfield, K. C., Coughlin, C. R., Friederich, M. W., Gallagher, R. C., Hesselberth, J. R., Lovell, M. A., Ofman, R., Swanson, M. A., Thomas, J. A., Wanders, R. J. A., Wartchow, E. P., and Van Hove, J. L. K. (2015) Mitochondrial energy failure in HSD10 disease is due to defective mtDNA transcript processing. *Mitochondrion*. **21**, 1–10

18. Deutschmann, A. J., Amberger, A., Zavadil, C., Steinbeisser, H., Mayr, J. A., Feichtinger, R. G., Oerum, S., Yue, W. W., and Zschocke, J. (2014) Mutation or knock-down of 17β-hydroxysteroid dehydrogenase type 10 cause loss of MRPP1 and impaired processing of mitochondrial heavy strand transcripts. *Hum Mol Genet*. **23**, 3618–3628

19. Falk, M. J., Gai, X., Shigematsu, M., Vilardo, E., Takase, R., McCormick, E., Christian, T., Place, E., Pierce, E. A., Consugar, M., Gamper, H. B., Rossmanith, W., and Hou, Y.-M. (2016) A novel HSD17B10 mutation impairing the activities of the mitochondrial RNase P complex causes X-linked intractable epilepsy and neurodevelopmental regression. *RNA Biol*. **13**, 477–485

20. Li, R., and Guan, M.-X. (2010) Human mitochondrial leucyl-tRNA synthetase corrects mitochondrial dysfunctions due to the tRNA$^{Leu(UUR)}$ A3243G mutation, associated with mitochondrial encephalomyopathy, lactic acidosis, and stroke-like symptoms and diabetes. *Mol Cell Biol*. **30**, 2147–2154

21. Oerum, S., Roovers, M., Leichsenring, M., Acquaviva-Bourdain, C., Beermann, F., Gemperle-Britschgi, C., Fouilhoux, A., Korwitz-Reichelt, A., Bailey, H. J., Droogmans, L., Oppermann, U., Sass, J. O., and Yue, W. W. (2017) Novel patient missense mutations in the HSD17B10 gene affect dehydrogenase and mitochondrial tRNA modification functions of the encoded protein. *Biochim Biophys Acta Mol Basis Dis*. **1863**, 3294–3302

22. Rossmanith, W., and Karwan, R. M. (1998) Impairment of tRNA processing by point mutations in mitochondrial tRNA[Leu(UUR)] associated with mitochondrial diseases. *FEBS Lett*. **433**, 269–274

23. Vilardo, E., and Rossmanith, W. (2015) Molecular insights into HSD10 disease: impact of SDR5C1 mutations on the human mitochondrial RNase P complex. *Nucleic Acids Research*. **43**, 5112–5119

24. Wang, S., Li, R., Fettermann, A., Li, Z., Qian, Y., Liu, Y., Wang, X., Zhou, A., Mo, J. Q., Yang, L., Jiang, P., Taschner, A., Rossmanith, W., and Guan, M.-X. (2011) Maternally inherited essential hypertension is associated with the novel 4263A>G mutation in the mitochondrial tRNAIle gene in a large Han Chinese family. *Circ Res*. **108**, 862–870

25. Metodiev, M. D., Thompson, K., Alston, C. L., Morris, A. A. M., He, L., Assouline, Z., Rio, M., Bahi-Buisson, N., Pyle, A., Griffin, H., Siira, S., Filipovska, A., Munnich, A., Chinnery, P. F., McFarland, R., Rötig, A., and Taylor, R. W. (2016) Recessive Mutations in TRMT10C Cause Defects in Mitochondrial RNA Processing and Multiple Respiratory Chain Deficiencies. *Am J Hum Genet*. **98**, 993–1000

26. Suzuki, T., Yashiro, Y., Kikuchi, I., Ishigami, Y., Saito, H., Matsuzawa, I., Okada, S., Mito, M., Iwasaki, S., Ma, D., Zhao, X., Asano, K., Lin, H., Kirino, Y., Sakaguchi, Y., and Suzuki, T. (2020) Complete chemical structures of human mitochondrial tRNAs. *Nature Communications*. **11**, 4269

27. Karasik, A., Wilhelm, C. A., Fierke, C. A., and Koutmos, M. (2021) Disease-associated mutations in mitochondrial precursor tRNAs affect binding, m1R9 methylation, and tRNA processing by mtRNase P. *RNA*. **27**, 420–432

28. Gobert, A., Gutmann, B., Taschner, A., Gößringer, M., Holzmann, J., Hartmann, R. K., Rossmanith, W., and Giegé, P. (2010) A single Arabidopsis organellar protein has RNase P activity. *Nat Struct Mol Biol*. **17**, 740–744

29. Gobert, A., Quan, Y., Arrivé, M., Waltz, F., Da Silva, N., Jomat, L., Cohen, M., Jupin, I., and Giegé, P. (2021) Towards plant resistance to viruses using protein-only RNase P. *Nat Commun*. **12**, 1007

30. Dreher, T. W. (2009) Role of tRNA-like structures in controlling plant virus replication. *Virus research*. **139**, 217

31. Reinhard, L., Sridhara, S., and Hällberg, B. M. (2015) Structure of the nuclease subunit of human mitochondrial RNase P. *Nucleic Acids Res*. **43**, 5664–5672

32. Nakamura, T., Yagi, Y., and Kobayashi, K. (2012) Mechanistic Insight into Pentatricopeptide Repeat Proteins as Sequence-Specific RNA-Binding Proteins for Organellar RNAs in Plants. *Plant and Cell Physiology*. **53**, 1171–1179

33. Howard, M. J., Lim, W. H., Fierke, C. A., and Koutmos, M. (2012) Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for

precursor-tRNA 5′ processing. *Proceedings of the National Academy of Sciences*. **109**, 16149–16154

34. Li, F., Liu, X., Zhou, W., Yang, X., and Shen, Y. (2015) Auto-inhibitory Mechanism of the Human Mitochondrial RNase P Protein Complex. *Sci Rep*. **5**, 9878

35. Teramoto, T., Kaitany, K. J., Kakuta, Y., Kimura, M., Fierke, C. A., and Hall, T. M. T. (2020) Pentatricopeptide repeats of protein-only RNase P use a distinct mode to recognize conserved bases and structural elements of pre-tRNA. *Nucleic Acids Res*. **48**, 11815–11826

36. Howard, M. J., Lim, W. H., Fierke, C. A., and Koutmos, M. (2012) Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5′ processing. *Proceedings of the National Academy of Sciences*. **109**, 16149–16154

37. Gobert, A., Bruggeman, M., and Giegé, P. (2019) Involvement of PIN-like domain nucleases in tRNA processing and translation regulation. *IUBMB Life*. **71**, 1117–1125

38. Gobert, A., Pinker, F., Fuchsbauer, O., Gutmann, B., Boutin, R., Roblin, P., Sauter, C., and Giegé, P. (2013) Structural insights into protein-only RNase P complexed with tRNA. *Nat Commun*. **4**, 1353

39. Reiter, N. J., Osterman, A., Torres-Larios, A., Swinger, K. K., Pan, T., and Mondragón, A. (2010) Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature*. **468**, 784–789

40. Bhatta, A., Dienemann, C., Cramer, P., and Hillen, H. S. (2021) Structural basis of RNA processing by human mitochondrial RNase P. *Nat Struct Mol Biol*. **28**, 713–723

41. Vilardo, E., Nachbagauer, C., Buzet, A., Taschner, A., Holzmann, J., and Rossmanith, W. (2012) A subcomplex of human mitochondrial RNase P is a bifunctional methyltransferase--extensive moonlighting in mitochondrial tRNA biogenesis. *Nucleic Acids Res*. **40**, 11583–11593

42. Karasik, A., Fierke, C. A., and Koutmos, M. (2019) Interplay between substrate recognition, 5′ end tRNA processing and methylation activity of human mitochondrial RNase P. *RNA*. **25**, 1646–1660

43. Reinhard, L., Sridhara, S., and Hällberg, B. M. (2017) The MRPP1/MRPP2 complex is a tRNA-maturation platform in human mitochondria. *Nucleic Acids Research*. **45**, 12469–12480

44. Zhang, J., and Ferré-DAmaré, A. R. (2016) Trying on tRNA for Size: RNase P and the T-box Riboswitch as Molecular Rulers. *Biomolecules*. **6**, 18

45. Howard, M. J., Klemm, B. P., and Fierke, C. A. (2015) Mechanistic Studies Reveal Similar Catalytic Strategies for Phosphodiester Bond Hydrolysis by Protein-only and RNA-dependent Ribonuclease P. *J Biol Chem*. **290**, 13454–13464

46. Howard, M. J., Karasik, A., Klemm, B. P., Mei, C., Shanmuganathan, A., Fierke, C. A., and Koutmos, M. (2016) Differential substrate recognition by isozymes of plant protein-only Ribonuclease P. *RNA*. **22**, 782–792

47. Chen, T.-H., Sotomayor, M., and Gopalan, V. (2019) Biochemical Studies Provide Insights into the Necessity for Multiple *Arabidopsis thaliana* Protein-Only RNase P Isoenzymes. *Journal of Molecular Biology*. **431**, 615–624

48. Vilardo, E., Toth, U., Hazisllari, E., Hartmann, R. K., and Rossmanith, W. (2023) Cleavage kinetics of human mitochondrial RNase P and contribution of its non-nuclease subunits. *Nucleic Acids Research*. 10.1093/nar/gkad713

49. Swanson, R. V. (2001) Genome of *Aquifex aeolicus*. in *Methods in Enzymology*, pp. 158–169, Hyperthermophilic Enzymes Part A, Academic Press, **330**, 158–169

50. Marszalkowski, M., Willkomm, D. K., and Hartmann, R. K. (2008) 5′-End maturation of tRNA in *Aquifex aeolicus*. *Biol. Chem.* **389**, 395–403

51. Willkomm, D. K., Feltens, R., and Hartmann, R. K. (2002) tRNA maturation in *Aquifex aeolicu*s. *Biochimie*. **84**, 713–722

Chapter 2 Structure Characterization of HARPs

2.1 Introduction

Transfer RNAs (tRNAs) play a key role in biology, shuttling amino acids into the ribosome during protein synthesis. Like other RNAs in the translation machinery, tRNAs are transcribed as precursors that must be matured before they attain their biologically active form. Precursor tRNA (pre-tRNA) maturation often includes the removal of 5' leader and 3' trailer sequences and/or introns in the anticodon stem loop, as well as the addition of nucleoside modifications (**Figure 2.1*A***). The removal of 5' leader sequences by ribonuclease P (RNase P) is one of the most highly conserved steps in tRNA maturation across all domains of life (1, 2). RNase P is an endoribonuclease that is essential in all organisms, with the exception of *N. equitans*, which does not encode 5' leader sequences in its tRNA (3–6).

RNase P has evolved independently multiple times, leading to a diverse set of RNase P molecular architectures across different species (**Figure 2.1*B***) (7). The ribonucleoprotein (RNP) form of RNase P was one of the first ribozymes discovered and has been studied for decades. This molecule consists of a catalytic RNA core associated with a varying number of accessory proteins (2, 8–10). However, RNase P is not always a ribozyme; proteinaceous, RNA-free forms of RNase P (termed protein-only RNase P, or PRORP) exist in all eukaryotes except Amoebozoa (11–13). PRORPs are single proteins with a catalytic metallonuclease domain, a central linker domain, and a pentatricopeptide repeat (PPR) substrate recognition domain (14). In metazoans, the mitochondrial form of the enzyme requires two additional protein partners to function *in vivo* (11).
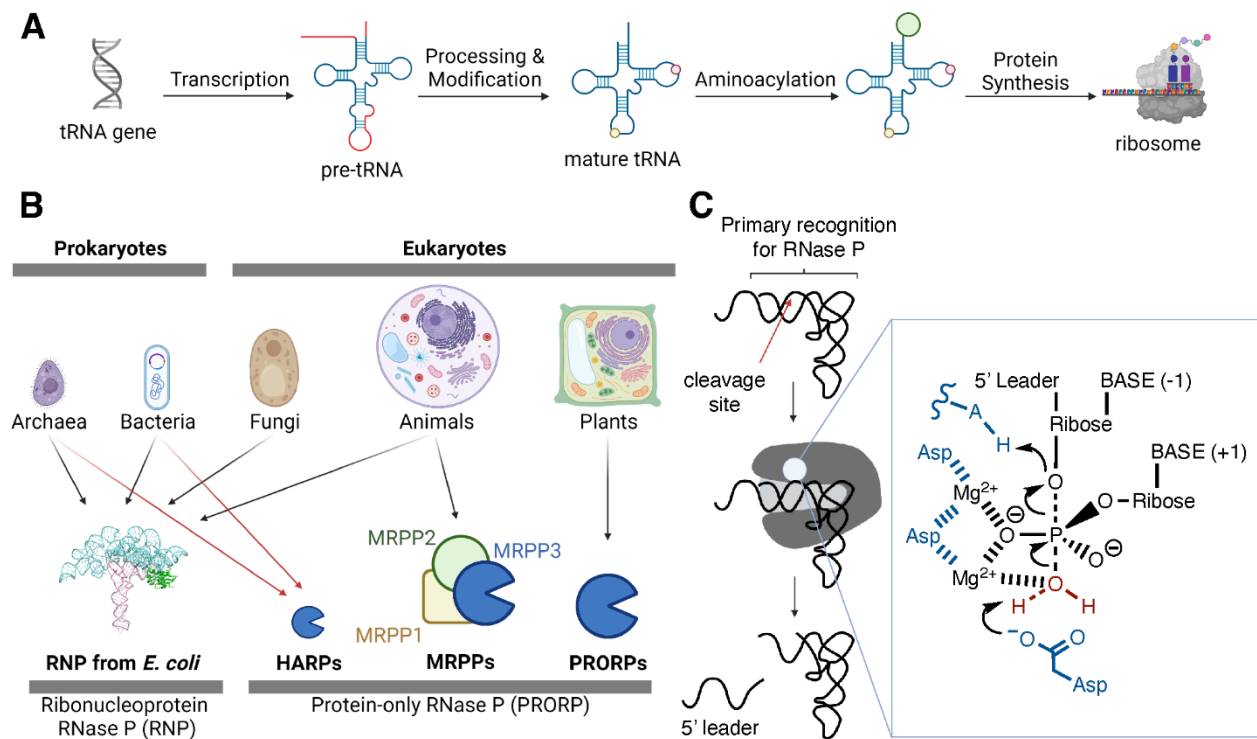
Figure 2.1: An overview of RNase P.

A. tRNA transcripts undergo processing, modification, and aminoacylation prior to use in translation. **B.** RNase P exists in all domains of life in different forms, including ribonucleoprotein complexes (RNPs) and protein only RNase P enzymes (PRORPs). PRORPs can be further categorized into the monomeric PRORP found in plants, the PRORP complex found in mitochondria (MRPP), and the HARP homo-oligomer found in bacteria and archaea. **C.** The conserved RNase P mechanism includes substrate recognition based on the 42 Å distance between the tRNA elbow and 5' cleavage site followed by divalent cation-mediated hydrolysis of the 3'-5' phosphate bond between the nucleotides at position -1 and +1.

Despite these major structural differences, RNase P enzymes have a conserved catalytic strategy. Catalysis is carried out by a divalent magnesium ion that hydrolyzes the phosphodiester bond between nucleotides at positions -1 and +1 of pre-tRNAs. This results in the formation of a tRNA with a mature 5' end with a phosphate group and a leader sequence with a 3' hydroxyl group (**Figure 2.1C**). In all RNP and PRORP RNase P enzymes characterized thus far, the active site is approximately 42 Å from the binding region. In PRORP, the binding region consists of positively charged amino acids that bind to the negatively charged phosphate backbone of the elbow region of the substrate pre-tRNA (15). This 42 Å distance is attributed to being a "molecular ruler" for all RNP and PRORP RNase P that measure from the T and D loops in the elbow region of the pre-tRNA to the cleavage site between the +1 and -1 nucleotides, a distance which is conserved among tRNAs (16).

Recently, a distinct class of protein-only RNase Ps were discovered in the thermophilic bacterium *Aquifex aeolicus* (17). The RNase P that was discovered in *A. aeolicus* is vastly different from previously characterized PRORPs. In contrast to other PRORPs, the *A. aeolicus* enzyme is quite small (23 kDa) and only shares sequence similarity with a single region of other PRORPs (17), the metallonuclease domain. Additional homologs of this small RNase P enzyme were identified in numerous bacterial and archaeal species, and these proteins are collectively referred to as <u>H</u>omologs of <u>A</u>quifex <u>R</u>Nase <u>P</u> (HARPs) (17).

The lack of an apparent substrate recognition domain in HARPs raised questions about how HARP enzymes interact with their pre-tRNA substrates. Initial research established that HARP from *A. aeolicus* exists as a large homo-oligomeric complex (17). To understand how this oligomerization may contribute to substrate recognition, the cryo-EM structure of a homo-dodecameric (12-mer) HARP assembly has been published by three independent groups. These

12-mer structures include HARP from *A. aeolicus* (Aq880) (18), *Halorhodospira halophila* SL1 (Hhal2243) (19), and *Thermocrinis ruber* (20). HARP structures in additional oligomeric states have also been published, including HARP dimers from *Planctomyces bacterium* and *Thermococcus celer* and a HARP tetramer from *P. bacterium* complexed with *E. coli* pre-tRNA[His] (20). These data all show that a monomer of HARP is a VapC-like endonuclease that consists of a PIN domain-like metallonuclease domain and two protruding α helices, which are thought to be responsible for substrate recognition (18, 19). The active site is conserved among HARPs and PRORPs, with five negatively charged residues, predominantly aspartate, positioned to coordinate two catalytic magnesium ions. The exact role of HARP enzymes *in vivo* remains to be elucidated, as most organisms that encode HARP also encode the RNP RNase P, and deletion of HARP in *M. mazei* and *H. volcanii* had no effect on their growth compared to the wild type strains (21).

The work presented here uses a combination of cryo-EM and X-ray crystallography to characterize three homo-oligomeric states of HARP: a tetramer, a dodecamer, and a novel tetradecamer (14-mer).

2.2 Materials and Methods

2.2.1 Expression and purification of Aq880 and Hth1307.

The HARP genes from *Aquifex aeolicus* (Aq880) and *Hydrogenobacter thermophilus* (Hth1307) were synthesized by Integrated DNA Technologies and amplified using PCR. Ligation-independent cloning was used to insert the genes into a pMCSG7 vector (22) that adds an N-terminal His$_6$-tag followed by a TEV cleavage site. The gene sequence was confirmed using Sanger sequencing through GeneWiz. The resulting plasmid was transformed and expressed in BL21 Rosetta DE3 *Escherichia coli* cells. Cells were grown at 37 °C to an OD$_{600}$ of 0.4 in TB medium and induced by the addition of 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG)

followed by growth for 18 hours at 22 °C. Harvested cells were resuspended in Lysis Buffer (20 mM MOPS pH 7.8, 1 M NaCl, 15 mM imidazole, 0.25% Tween-20, 1 mM TCEP) with the addition of the following per 100 mL: $MgCl_2$ to 2.5 mM, 10 mg of lysozyme, 3.5 µL of 250 units/µL Benzonase, phenylmethylsulfonyl fluoride (PMSF) to 250 µM, and two cOmplete EDTA-free protease inhibitor cocktail mini-tablets (Roche Applied Science). Cells were lysed using a Q500 sonicator (Qsonica) at 60% amplitude with pulses of 5 seconds on/10 seconds off for a total of 10 minutes on. The lysate was centrifuged for 30 minutes at 18,000 rpm at 4 °C with a JA-20 rotor (Beckman Coulter). The soluble fraction was applied to a Ni-Sepharose HisTrap HP 5 mL column (Cytiva Life Sciences) pre-equilibrated with Lysis Buffer. The column was washed with 5 column volumes (CV) of Lysis Buffer and 5 CV of Nickel Wash Buffer (20 mM MOPS pH 7.8, 150 mM NaCl, 50 mM imidazole, 1 mM TCEP). Bound proteins were eluted by a gradient from 50 to 500 mM imidazole over 20 CV. Fractions containing HARP were pooled and dialyzed against DEAE Buffer (50 mM Tris pH 8.5, 150 mM NaCl, 1 mM TCEP) overnight at 4 °C. The dialyzed sample was loaded onto a DEAE Sepharose FF column (Cytiva Life Sciences) and eluted with a 150 to 1000 mM NaCl gradient over 20 CV. Fractions containing HARP were pooled and 1 mg tobacco etch virus protease per 50 mg HARP was added to the sample and dialyzed against SEC Buffer (50 mM Tris pH 8.5, 150 mM NaCl, 1 mM TCEP) overnight at 4 °C with constant stirring. The sample was then applied to a second Ni-Sepharose column. The flow-through was collected and concentrated with a 10 kDa MWCO Amicon Ultra-15 spin column. The sample was then loaded onto a Superdex 200 pg 16/600 (Cytiva Life Sciences) size exclusion column and eluted with SEC Buffer. Peak fractions were pooled and concentrated with a 10 kDa MWCO Amicon Ultra-15 spin column before using in experiments or flash freezing for storage at -80 °C.

2.2.2 Analytical SEC to determine concentration dependent oligomerization of Hth1307.

Samples of Hth1307 were diluted in SEC Buffer to concentrations ranging from 93.6 nM to 144 µM and filtered with 0.22 µm Costar Spin-X centrifuge tube filters. Samples of 150 µL were loaded onto a Superdex 200 Increase 5/150 GL analytical size exclusion column (Cytiva Life Sciences) and eluted at 0.17 mL/min while monitoring the absorbance at 280 nm. Elution profiles were overlaid using Prism 9.

2.2.3 Native mass spectrometry of Hth1307.

Native mass spectra were collected on a ThermoFisher Q-Exactive Ultra-High Mass Range (UHMR) Hybrid Quadrupole-Orbitrap Mass Spectrometer in a positive ionization polarity. Samples were buffer exchanged into 200 mM ammonium acetate (AmOAc) pH 7.0 using Micro Bio-Spin P-6 gel columns and directly infused via nanoelectrospray ionization (nESI). nESI was performed using borosilicate needles pulled and coated in-house with a Sutter p-97 Needle Puller and a Quorum SCX7620 mini sputter coater, respectively. Mass spectrometry parameters of note include a capillary temperature between 200-250 °C, the high-collisional dissociation (HCD) cell set between 10-20 V, and the spray voltage set between 1.0-1.4 kV. Detector optimization and transfer optics were set to high m/z and spectra were collected at 1 minute. The acquired native mass spectra were deconvoluted using UniDec (23).

2.2.4 Crystallization of Hth1307 and data collection.

Purified protein was concentrated using ultrafiltration (Centricon-30, Amicon) to 10 mg/mL. Crystals of Hth1307 were grown using the sitting drop vapor diffusion method. Crystallization of the purified Hth1307 was initially performed with commercially available screens at 20 °C and 4 °C. Each experiment consisted of mixing 0.5 µL of protein solution (10 mg/mL Hth1307 in 50 mM Tris-HCl pH 8.5, 150 mM NaCl, 1 mM TCEP; Protein Sample) with 0.5 µL of reservoir solution and equilibrate the drop against 50 µL of reservoir solution in

INTELLI-PLATE 96-well trays (Art Robbins). Crystals of the Hth1307 in the tetramer form were grown at 4 °C by mixing Protein Sample with a crystallization solution consisting of 2 M ammonium sulfate and 0.1 M Bis-Tris pH 5.5. Crystals of the Hth1307 tetramer were briefly transferred to a cryo-protectant solution containing 20% (v/v) glycerol, 1.6 M ammonium sulfate, and 0.08 M Bis-Tris pH 5.5 before harvesting and flash-freezing in liquid nitrogen. Hth1307 crystals in the 14-mer form were grown at 4 °C by mixing Protein Sample with a crystallization solution containing 1.0 M sodium malonate pH 5.0, 0.1 M sodium acetate trihydrate pH 4.5, and 2% (w/v) polyethylene glycol 20,000. Crystals were briefly transferred to a cryo-protectant solution containing 20% (v/v) glycerol, 0.8 M sodium malonate pH 5.0, 0.08 M sodium acetate trihydrate pH 4.5, and 1.6 % (w/v) polyethylene glycol 20,000 before harvesting and flash-freezing in liquid nitrogen.

Diffraction data were collected at the GM/CA 23-ID-B beamline at the Advanced Photon Source (APS), Argonne National Laboratory using a Dectris Eiger X 16 M detector. Data collection and processing statistics are summarized in **SI Table 2.1**. Data for the Hth1307 tetramer were indexed to space group $C222_1$ (unit-cell parameters a = 90.4, b = 108.9, c = 105.7 Å) with two protomers in the asymmetric unit (Matthew's coefficient V_M = 2.92 $Å^3$ Da^{-1}, 58% solvent content). Data for the Hth1307 14-mer were indexed to space group $P2_122_1$ (unit-cell parameters a = 102.3, b = 113.8, c = 155.3 Å) with seven protomers in the asymmetric unit (Matthew's coefficient V_M = 2.88 $Å^3$ Da^{-1}, 57% solvent content).

2.2.5 Structure solution of Hth1307 crystallography data.

Data sets for the HARP tetramer and tetradecamer crystals were processed using xia2/DIALS (24). Initial phases for the tetramer HARP structure were obtained through molecular replacement with Phaser (25) using a single monomer from *Halorhodospira halophila* HARP

(7OG5 (19)) as a search model. Initial phases for the tetradecameric HARP structure were obtained through molecular replacement with Phaser (25) using a monomer from our tetrameric HARP structure (8SSF). Following molecular replacement, iterative model building and corrections were performed manually using Coot (26). Initial refinement for both structures was conducted using BUSTER version 2.10.4 (27) to rapidly fix Ramachandran, rotamer, and density fit outliers, refining to convergence and adding waters in the final automated round of refinement. Subsequent structure refinements were performed using CCP4 Refmac5 (28) with non-crystallographic symmetry (NCS) for the tetradecameric structure and without NCS for the tetrameric structure, including restrained refinement, coordination minimization, and restrained individual B-factor adjustment with maximum-likelihood targets. PDB-REDO (29) was used to assess the model quality between refinements and to fix any rotamer and density fit outliers automatically. The model quality was evaluated using MolProbity (30). Figures showing crystal structures were generated in PyMOL (31). Structural alignments were performed using the extra_fit function in PyMOL. Angles between tetramers were calculated using anglebetweendomains.py in PyMOL. Atomic coordinates and structure factors for the two crystal structures have been deposited in the Protein Data Bank (PDB).

2.2.6 Cryo-EM grid preparation, data acquisition and data processing of Aq880.

Freshly purified Aq880 was screened first using 1% uranyl formate in Morgagni operated at 100kv and images collected at a magnification of 22,000x. For cryo-EM, 3 µl of Aq880 sample at a concentration of 8 µM was applied to Lacey carbon 200 mesh grid (SPI supplies), which was glow discharged with EasiGlow glow discharge for 60s (setting 5 mA, 0.26 mBar). The grids were then flash frozen in liquid ethane using a Thermo Fisher Scientific Vitrobot IV (30 s wait time, 2.5 s blot time, -5 blot force) using Whatman filter paper 1. Cryo-EM images were collected on a

Thermo Fisher Scientific Glacios cryo-TEM operated at 200kv using a Gatan K2 Summit® camera using Leginon. The image was recorded at a dose rate of 7.7 e⁻/Å² with 200ms exposure time and collected 40 frames. Defocus ranges ranged from -1.5 to -2.0 μm. Data was processed using cryoSPARC, following the workflow of the cryoSPARC software (32). Motion correction (33, 34) and contrast-transfer function was performed using full-frame motion correction and path CTF estimation (34, 35). Particle picking was performed using template picker, picked particles were extracted and performed reference free 2D classification in cryoSPARC (32). After iterative rounds of 2D class average, good classes were used to build two ab-initio models using ab-initio reconstruction in cryoSPARC (32). Resulting maps were then used for heterogeneous refinement using the full set of good particles in cryoSPARC. Particles bearing the most homogeneous density were subjected to non-uniform 3D refinement yielding a resolution of 7.8Å volume. Resolution was estimated according to the Gold-Standard Fourier Shell (GS-FSC) = 0.143 criterion (36). Figures showing cryo-EM structures were generated using UCSF ChimeraX (37, 38).

2.3 Results

2.3.1 Aq880 and Hth1307 HARP occupy different oligomeric states.

We concurrently worked on Aq880 and Hth1307 HARPs, both of which expressed and purified as a mixture of oligomers. When purifying Aq880, we observed tetramers and dodecamers that were difficult to separate. Single particle analysis with cryo-EM was used to investigate the structure of Aq880 as it is uniquely suited to deconvoluting mixtures. Close inspection of the 2D class averages reveals orientation bias in our dodecamer particles (**Figure 2.2**). From this data, we were able to create a low-resolution model of the Aq880 dodecamer (**Figure 2.2**), which agrees with the radial arrangement of dimers seen in previously published structures (18–20).

29

Next we focused our attention on Hth1307, whose multiple oligomeric states were more easily separated during purification. The predominant product was a large oligomeric assembly, although there was also a small proportion of enzyme that purified as a smaller oligomeric assembly. These isolated fractions were used to successfully grow protein crystals of Hth1307 in two different oligomeric states, described below. The large Hth1307 assembly eluted sooner than that of Aq880 on the elution profile during size exclusion chromatography, indicative of a higher molecular weight and oligomeric state (**Figure 2.3**).

Native mass spectrometry (MS) with nano-electrospray ionization was used to quantify the oligomeric populations of Hth1307. This method separates proteins and protein complexes by size, shape, and charge prior to MS analysis, providing accurate and detailed mass measurements and stoichiometry information (39). Experiments using His-tagged Hth1307 revealed a single oligomeric state of a tetradecamer (14-mer, **Figure 2.4*A* and Figure 2.5**). While there is a difference in the observed and expected molecular weights of the 14-mer, these differences were identified to be a result of labile His-tags (40). Specifically, an average of 5-6 tags were lost during MS analysis. In the mass spectra, two charge envelopes of the 14-mer can be readily observed. While the high charge states could be representative of a more unfolded 14-mer due to coulombic repulsion (41–43), this bimodality potentially indicates two predominate conformational states that were previously hypothesized (18–20). Prior to this point, the highest oligomeric state of HARP proteins was shown to be a dodecamer (12-mer).
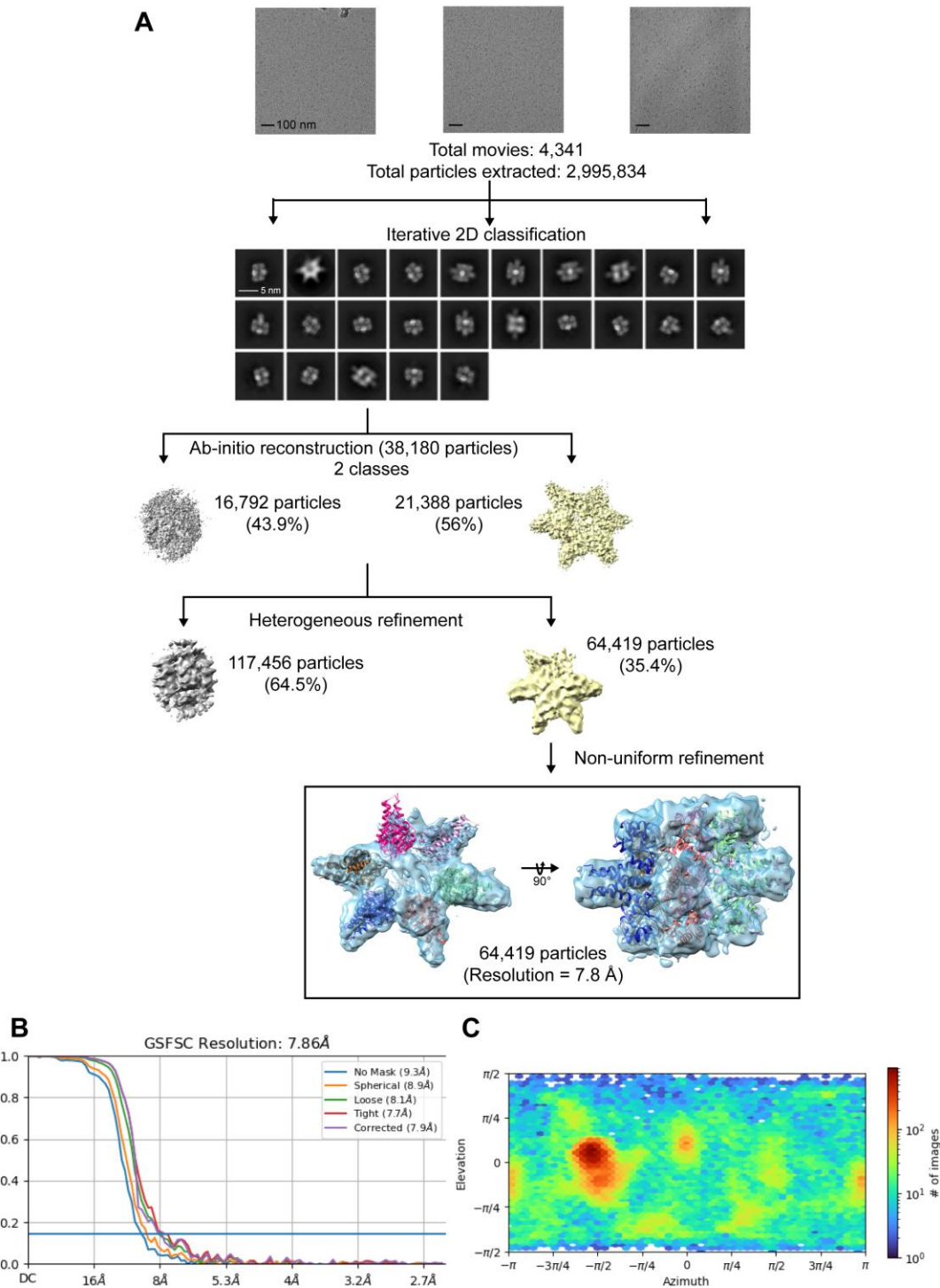
Figure 2.2: Cryo-EM data processing workflow.

A. Cryo-EM micrographs were collected with a Glacios microscope operating at 200 kV and a pixel size of 1.25 Å. Computational processing was performed using cryoSPARC v3.1. The black scale bar on the micrographs is 100 nm. The white scale bar on the class averages is 5 nm. Full details of data collection and processing can be found in the Methods section. **B.** Resolution estimated by Gold Standard-FSC = 0.143. **C.** Particle distribution shows evidence of orientation bias.
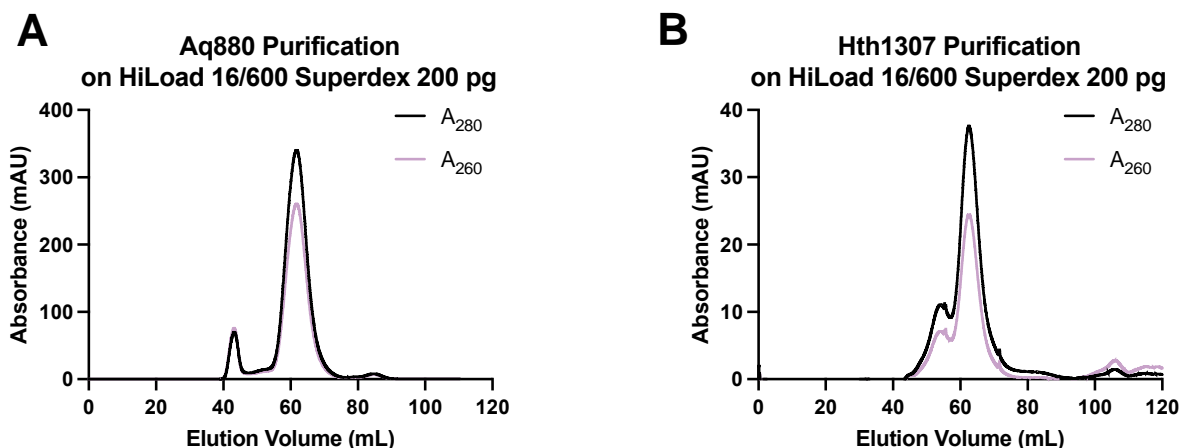
Figure 2.3: Elution profiles from Aq880 and Hth1307 protein purifications.

A-B. Representative elution profiles from the size exclusion chromatography with the HiLoad 16/600 Superdex 200 pg column for both Aq880 and Hth1307, showing that Hth1307 has a sub-population that elutes at a higher molecular weight. The small peak at 40 mL in **A.** is the elution of aggregated protein after the dead volume of the column.
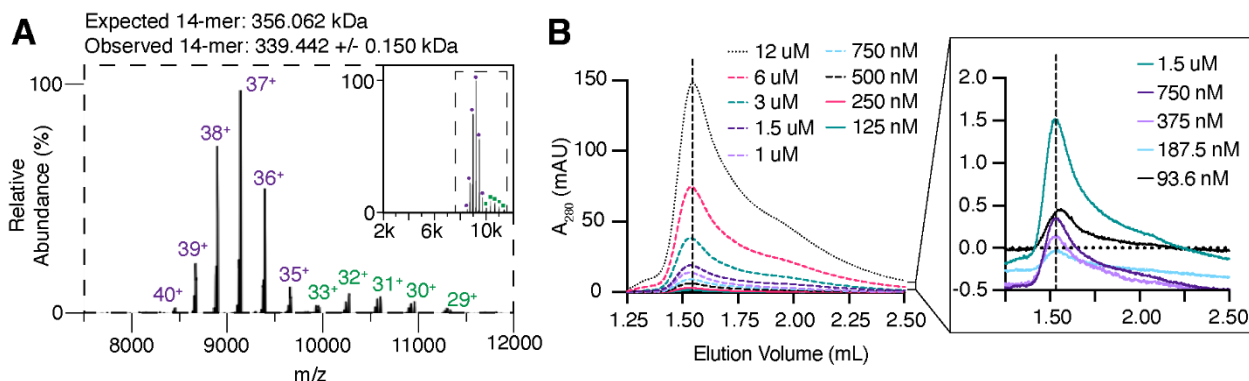
Figure 2.4: Native MS and SEC demonstrating oligomeric states of Hth1307

A. Native mass spectra were obtained using His-tagged Hth1307. No oligomeric states below or above a 14-mer are observed. Differences in the expected and observed molecular weights are due to partial loss of labile His-tags. Two charge state series (or charge envelopes) are observed for the 14-mer species (purple circles and green squares) indicative of two main conformational states. The high charge state series is shown in purple and the low charge state series is shown in green. A detailed and zoomed view of this data can be found in Figure S5. **B.** Size exclusion chromatography was conducted using Hth1307 samples ranging from 93.6 nM to 12 μM as indicated in the graph legend while monitoring the absorbance at 280 nm to look for changes in the molecular weight of the complex via the elution volume. A size exclusion profile using standard molecular weight proteins can be found in **SI Figure 2.2**. A260 levels lower than the A280 confirmed a lack of RNA contamination (**SI Figure 2.4**). The inset is zoomed in to a lower absorbance range.

A. Expected 14-mer: 356.062 kDa
Observed 14-mer: 339.442 ± 0.150 kDa

High Charge State Series
Low Charge State Series

B.
337.855 ± 0.035 kDa · 339.052 ± 0.014 kDa
338.585 ± 0.013 kDa

C.
338.579 ± 0.223 kDa · 340.154 ± 0.305 kDa
339.276 ± 0.306 kDa · 327.265 ± 0.049 kDa
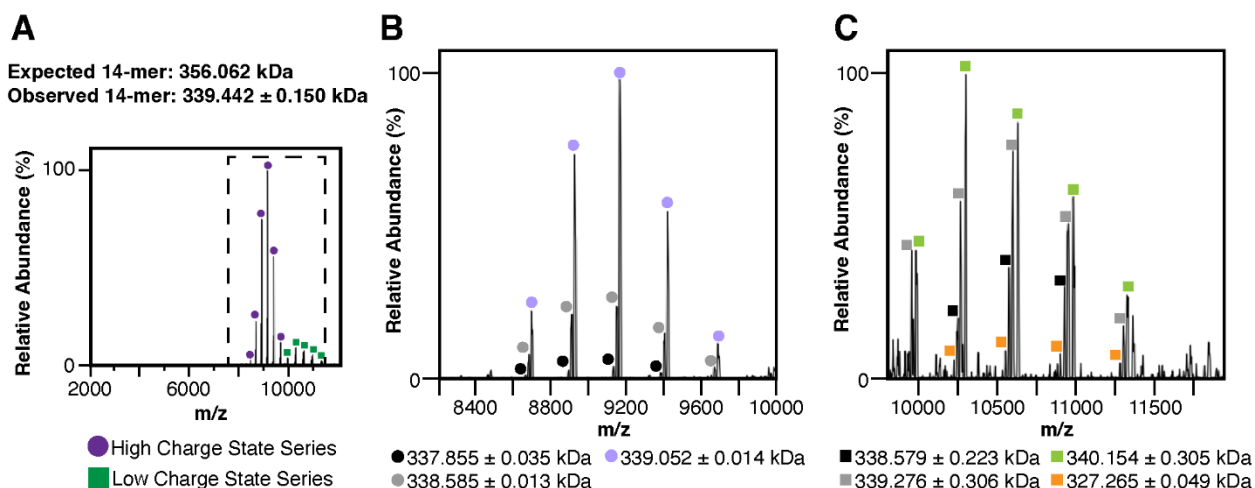
Figure 2.5: Detailed analysis of native mass spectra results for Hth1307.

A. Native mass spectra of His-tagged Hth1307. Upon a closer look at the **B.** high and **C.** low charge state series, additional masses are observed. In each of these series, masses differing by between 200-600 Da are observed and could be related to cleavage of a portion of the labile His-tag or loss of the magnesium ions from the active site. **C.** Within the low charge state series, a new series is observed (orange squares) and relates to a 14-mer with the loss of 9-10 His-tags.

Because native MS is limited to larger concentrations of protein (500-1000 nM), these experiments were accompanied by size exclusion chromatography experiments to monitor the elution volume, which corresponds to molecular weight, of Hth1307 at a variety of lower concentrations more similar to the assay conditions described below. There was no observed shift in the elution volume of the major population at 1.54 mL when starting with concentrations ranging from 94 nM to 126 μM (**Figure 2.4*B* and Figure 2.6**). Combined, these results demonstrate that while Hth1307 exists in many oligomeric states in solution, the majority of the sample is a tetradecamer (14-mer).

Figure 2.6 (next page)**:** Analytical SEC elution profiles for Hth1307.

A. Elution profiles of Hth1307 injected onto a Superdex 200 Increase 5/150 GL column at concentrations ranging from 250 nM to 12 μM, showcasing the ratio of the absorbance at 280 and 260 nm, corresponding to protein and RNA respectively. Lower sample concentrations did not yield reliable A260 measurements. **B.** Elution profile of Bio Rad Gel Filtration Standards showing the absorbance at 280 nm on the same column used in **A. C.** Standard curve from the elution profile on the left. This curve was used to approximate the molecular weight, and therefore oligomeric state, of Hth1307 at varying concentrations. Based on the average elution volume of 1.54 mL, the calculated molecular weight of the complex is 399.762 kDa, or approximately an 18-mer. However, larger molecular weights have more error in such standard curves due to the logarithmic relationship between elution volume and molecular weight. The oligomeric state of Hth1307 cannot be accurately determined by size exclusion chromatography but does not appear to change elution volume, and therefore oligomeric state, based on concentration.

A

12 uM HARP

6 uM HARP

3 uM HARP

1.5 uM HARP

1 uM HARP

750 nM HARP

500 nM HARP

250 nM HARP
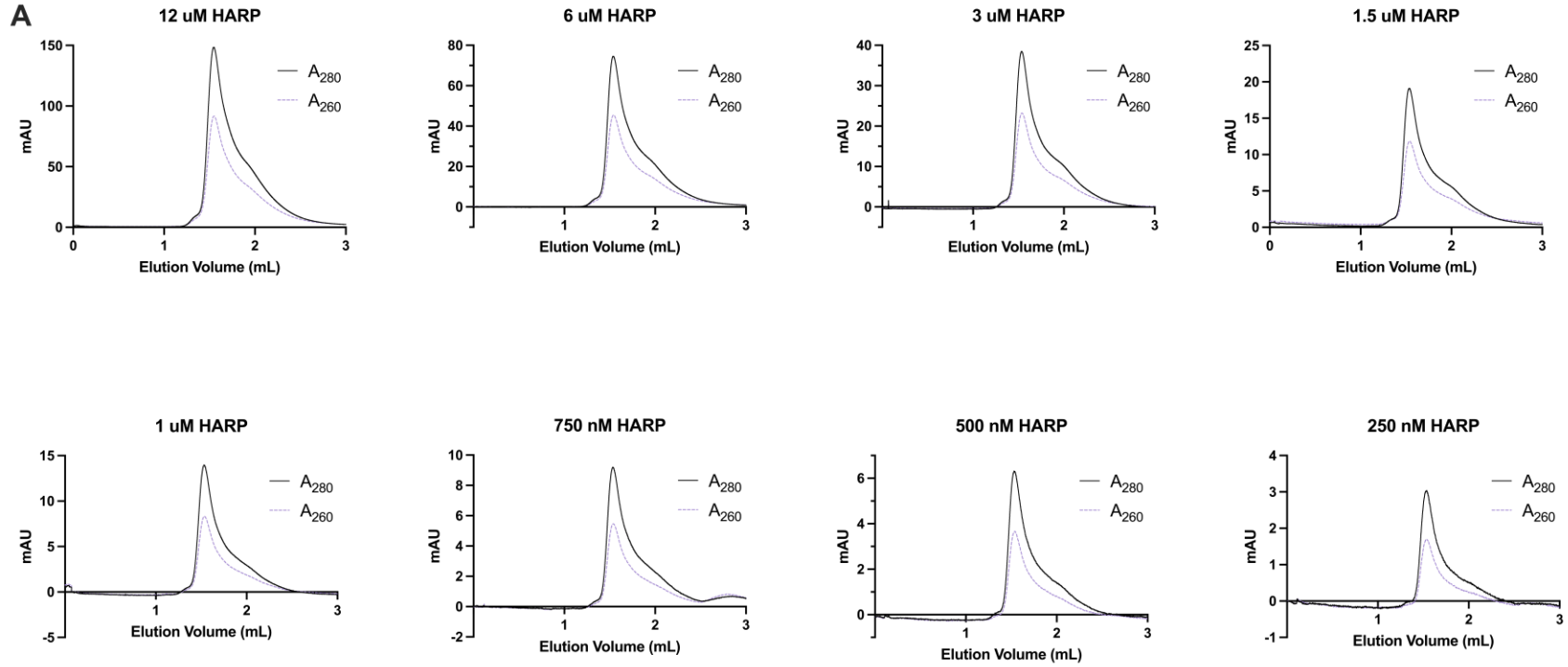
B Bio Rad Gel Filtration Standards on Superdex 200 Increase 5/150 GL

1. 1.35 mL (670,000 Da)
2. 1.78 mL (158,000 Da)
3. 2.08 mL (44,000 Da)
4. 2.34 mL (17,000 Da)
5. 2.73 mL (1,350 Da)
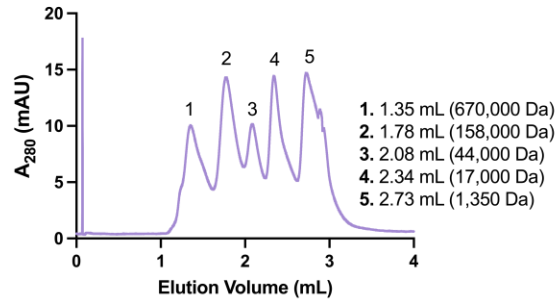
C Protein Standard Curve on Superdex 200 Increase 5/150 GL

$y = -1.914x + 8.542$
$R^2 = 0.9781$

$x = 1.52$ mL
$y = 5.63$
MW = 430,000
Oligomer = 18-mer

34
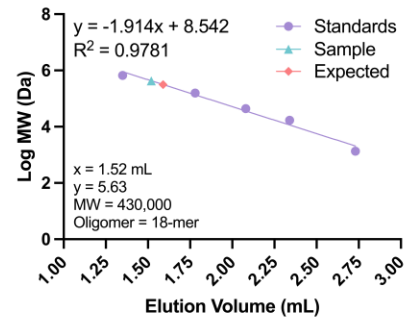
2.3.2 The X-ray structure of the Hth1307 tetramer reveals the minimal functional protein assembly in substrate-free form.

We sought to better understand the structure and function of Hth1307 using protein crystallography. From the lower molecular weight fraction of Hth1307 discussed above, we first determined the crystal structure of Hth1307 to 2.5 Å as a dimer within the asymmetric unit. The homotetrameric Hth1307 assembly is generated through symmetry by two identical dimers resulting in a dimer of dimers (**Figure 2.7*A-B*, SI Table 2.1**). Each monomer in the Hth1307 dimer consists of a PIN-like metallonuclease domain and a spike helix (SH) domain.

The metallonuclease domain consists of an α/β/α domain with a central four-stranded parallel β-sheet (**Figure 2.8*A***). Alignment of the Hth1307 metallonuclease domain with that of At PRORP1 (44) demonstrates a conserved active site with five negatively charged residues positioned to coordinate two divalent metal ions (**Figure 2.7*A*, inset**) (45). Although Hth1307 does not have a residue directly comparable to D497 in At PRORP1, D144 in Hth1307 is positioned in a manner that may fill a similar role upon conformational rearrangement. However, as of yet, there are no structures of HARP with a metal ion properly chelated for catalysis, nor have mechanistic studies established if HARP requires two metal ions for endonuclease activity as shown in *At* PRORP studies (45).

The SH domain is a protruding α-helical domain consisting of helices α6 and α7 (**Figure 2.7*A***), which substantially contribute to the dimerization interface by forming an extensive four-helical bundle. The dimer interface is further stabilized by interactions highlighted in **Figure 2.8*B*, Figure 2.9, and Figure 2.10,** and has a total surface area of 1459.3 $Å^2$ as calculated by PISA (46). In addition to the predominantly hydrophobic interactions, electrostatic interactions include hydrogen bonding between β4 of each monomer, hydrogen bonding between Y87 and D146' in

the SH domain, and salt bridges in the SH domain between residues E105 and R131' and residues

E105' and R127 (**Figure 2.8*B* and Figure 2.9**).
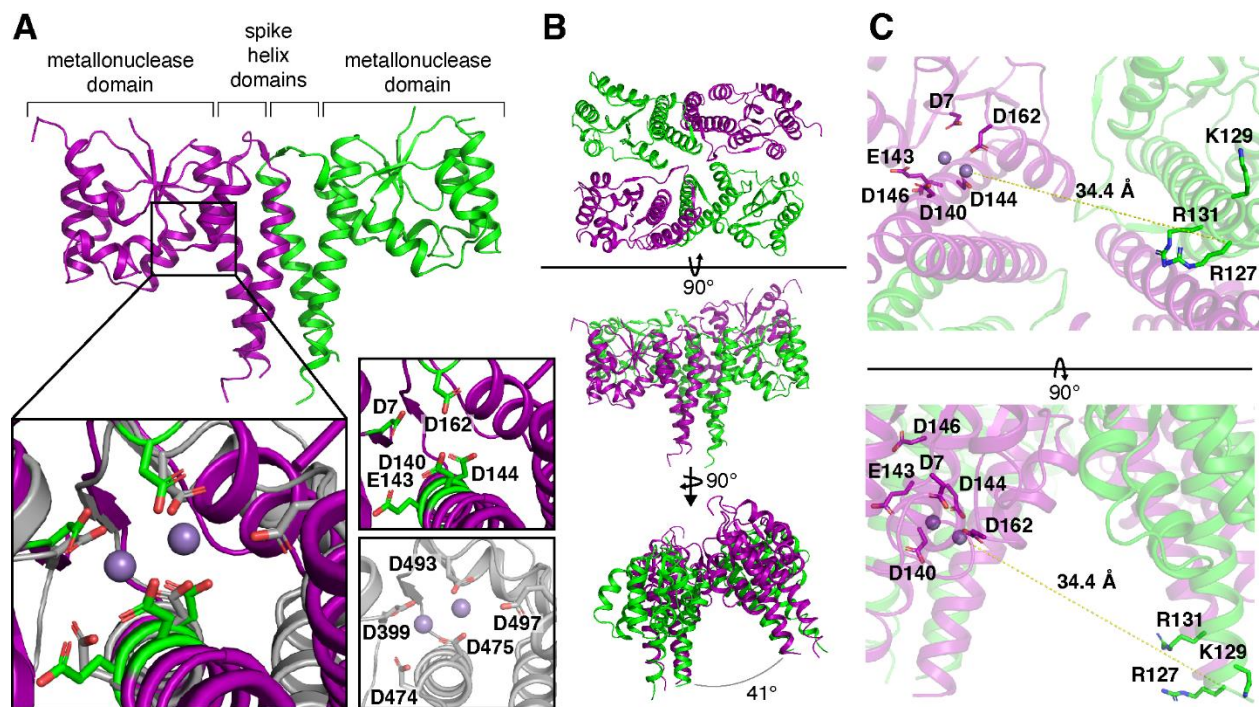
Figure 2.7: Crystal structure of the Hth1307 tetramer.

A. One monomer shown in purple and one in green. The inset zooms in on the active site of Hth1307 (active site residues shown in green) aligned to that of *At* PRORP1 (4G24, grey). The divalent cations (Mn^{2+}) from the *At* PRORP1 structure are shown in light purple. **B.** Generation of crystal symmetry mates for a Hth1307 dimer in the asymmetric unit reveals the captured structure is a tetramer, shown here in three views. **C.** The distance between the Mn^{2+} ions from *At* PRORP1 to the C_α of residue R127 of Hth1307 was used to approximate the distance between the active site of one monomer and the proposed binding site in the opposing monomer, which is 34.4 Å. The divalent cations were modeled based on the *At* PRORP1 structure.
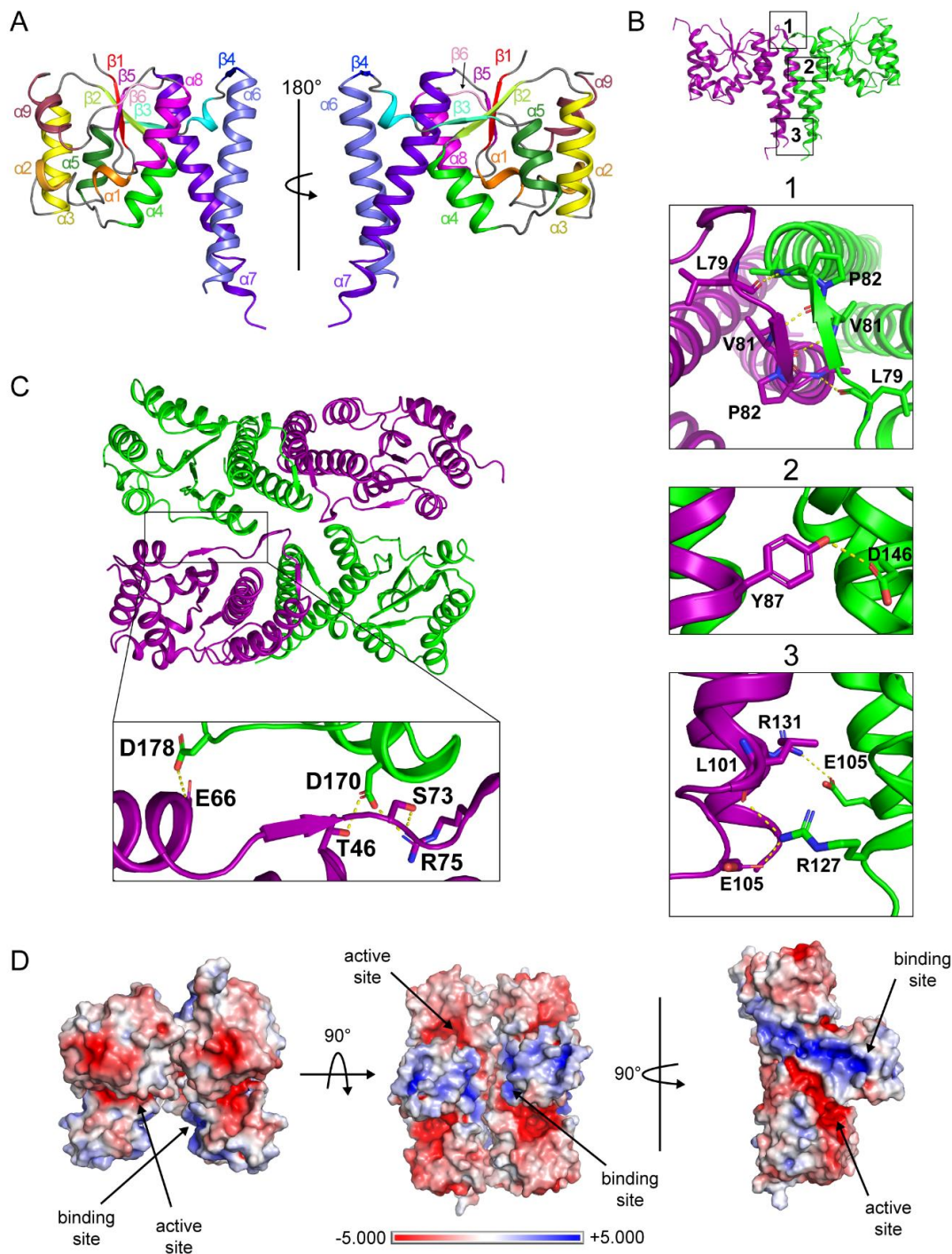
Figure 2.8: Details of the Hth1307 structure and interfaces.

A. An Hth1307 monomer with α helices and β strands numbered and color-coded. **B-C.** Dimer and tetramer interfaces in the Hth1307 tetramer. The residues involved in polar contacts are labeled with dashed yellow lines indicating hydrogen bonding and salt bridges. **B.** The dimer interface is a 4 helical bundle formed by the spike helix domains of both dimers. In addition to several hydrophobic interactions, the interface is stabilized by polar contacts in three areas. **C.** The tetramer interface forms between the metallonuclease domains of the monomers rather than the spike helix domains. **D.** Hth1307 surface shown with potential isocontours, calculated using the APBS Electrostatics plugin for PyMOL, with a range of +5 kT/e (blue) to -5 kT/e (red).
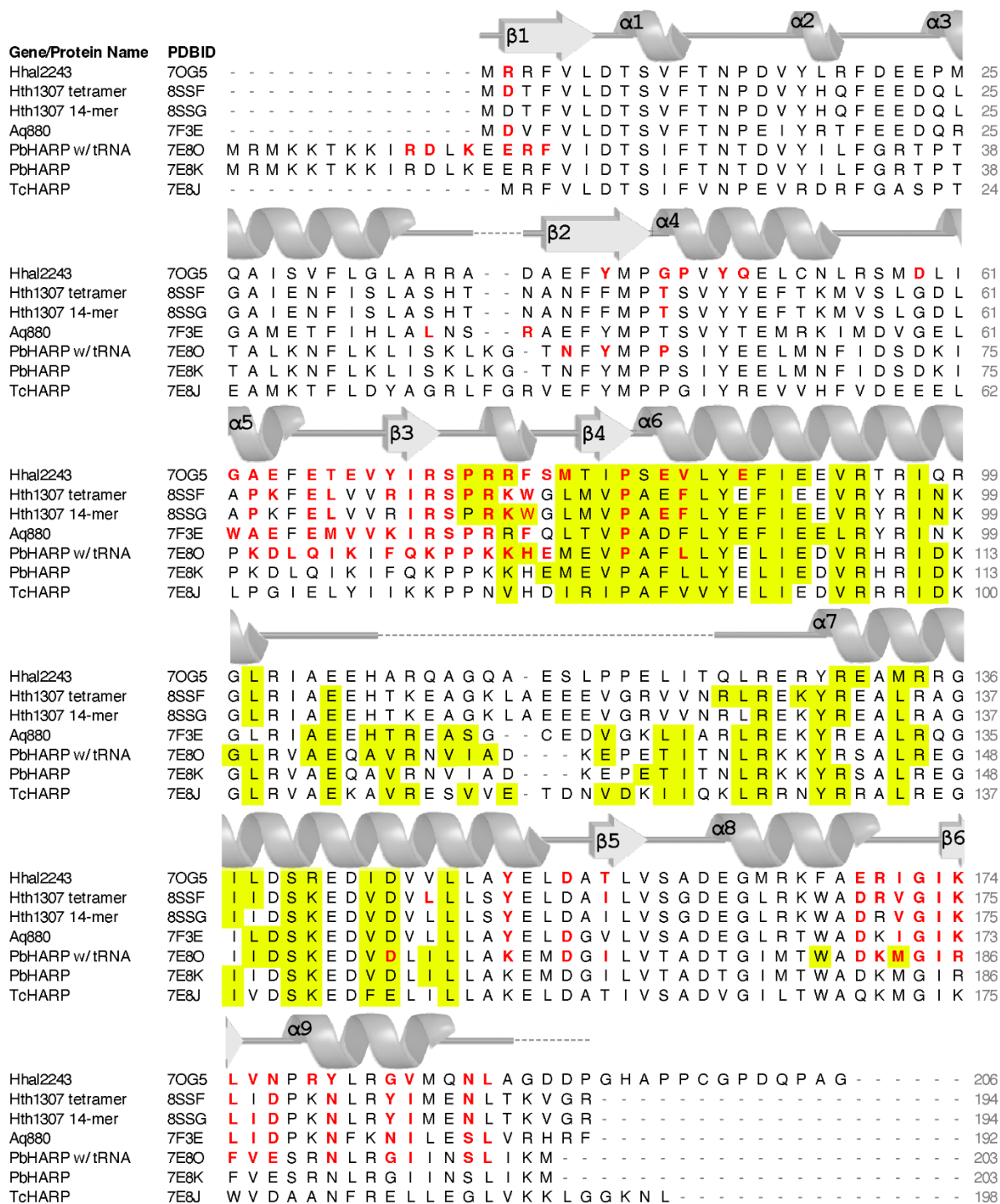
Figure 2.9: HARP sequence alignments.

Alignment of sequences of structurally characterized HARPs was performed using Clustal Omega. The secondary structure diagram above the alignments corresponds to the Hth1307 tetramer structure and was created using PISA. Residues highlighted in yellow or shown in red are part of the dimer and tetramer interface respectively.
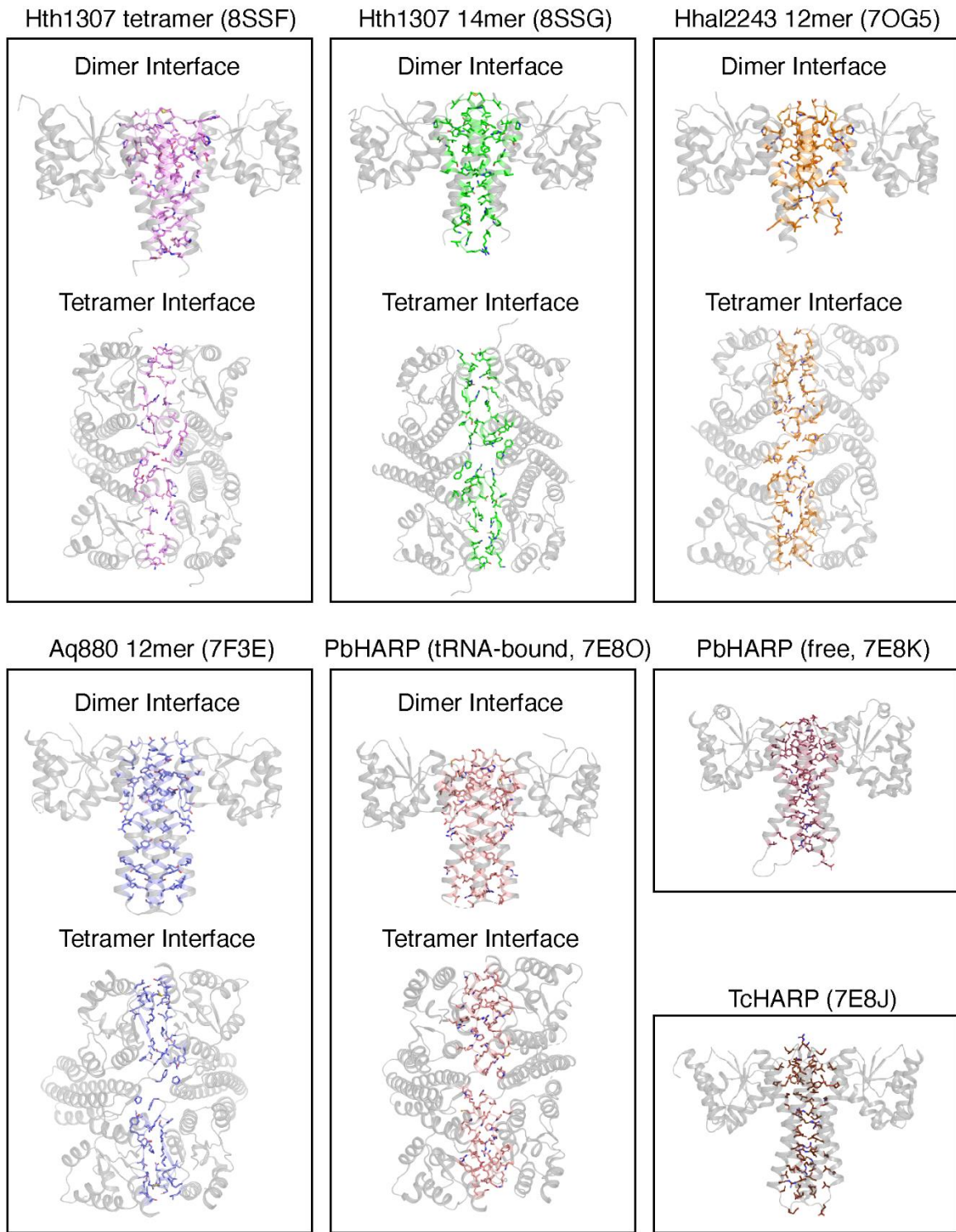
Figure 2.10: HARP interface comparisons.

Comparison HARP dimer and tetramer interfaces in structures from different species. The overall proteins are shown as cartoon in grey with the interface residues shown as sticks in color.

In the tetramer, two dimers come together in a side-by-side arrangement where the two dimers are angled at 41° (**Figure 2.7B**). The interface is mostly formed by contacts between the C-terminus of α5 through the N-terminus of α6 of one protomer (residues 63-77) and the C-terminus of α8 through the C-terminus of α9 in the opposing protomer (residues 170-189, **Figure 2.8CFigure 2.9**). The tetramer interface is considerably smaller than the dimer interface, with a surface area of 589.5 $Å^2$ between monomers, for a total of 1179.0 $Å^2$ across the entire interface (**SI Table 2.2**) (46). This interface has only two salt bridges, both between R75 and D170', although there are a number of other residues that participate in backbone and side chain hydrogen bonding, hydrophobic interactions, and pi-pi stacking (**Figure 2.8CFigure 2.9**). As seen in **Figure 2.9**, most of the residues involved in the dimer interface are conserved, likely to preserve the interface.

Tetramer formation has been proposed to be important for pre-tRNA binding (19, 20). Both ribonucleoprotein and protein-only RNase P enzymes recognize the tRNA elbow partly through shape complementarity. Both the RNPs and PRORPs form a large cleft spanning ~42 Å from the active site to an RNA-binding region, which is wide enough to accommodate the entire length of the tRNA acceptor stem (16). The 5' end of the acceptor stem is positioned in the metal active site, and the T and D loops are recognized by the pentatricopeptide repeat domain in PRORP (47–49) and MRPP3 (50) and the CR-II/CR-III modules in human and bacterial RNP RNase P (51, 52). This arrangement allows RNase Ps to recognize tRNA structural elements and select substrates through a molecular ruler mechanism (47, 51–55). Upon inspection of the electrostatic surface of the tetramer, there is a positively charged region of the SH domain (**Figure 2.8D**, (56)) that includes residues R125, R127, K129, and R131 (**Figure 2.7C**), which have been confirmed to be necessary for pre-tRNA processing activity (19). These residues are highly conserved across

HARPs from different species. In this structure, the positively charged region is 34.4 Å from the active site of the neighboring dimer (**Figure 2.7*C***), which is not wide enough to accommodate the canonical structure of tRNA without any large protein conformational rearrangement. Therefore, the conformation captured in this crystal structure is unlikely to be the conformation that supports substrate binding and nuclease activity. However, as discussed later in this paper, kinetic experiments demonstrate the protein we isolated maintains nuclease cleavage activity, possibly through substrate-induced conformational changes that have yet to be captured or characterized.

2.3.3 Structural characterization of the Hth1307 tetradecamer using crystallography.

The high molecular weight fraction of Hth1307 produced crystals under different conditions than those of the tetramer described above. The larger unit cell of these crystals implied a larger oligomeric formation. Molecular replacement revealed a heptamer in the asymmetric unit that, upon generation of its symmetry mate, produces a tetradecamer (14-mer) flat cylindrical ring consisting of a radial arrangement of seven HARP dimers (**Figure 2.11*A*, SI Table 2.1**). The structure was solved to 3.2 Å resolution. We aligned the seven unique monomers from the 14-mer to both of the unique monomers from our tetramer structure (**Figure 2.11*B***). This demonstrates that all of the metallonuclease domains in our Hth1307 tetramer and 14-mer structures are nearly identical, with RMSD values ranging from 0.87-1.16 Å over 1021-1088 atoms. Notably, the dimer and tetramer interfaces of the 14-mer are fairly similar when compared to our Hth1307 tetramer structure (**Figure 2.9**). The angles between dimers in the 14-mer structure range from 48° to 54°, and the distances between the proposed tRNA elbow binding region and the active site ranges from 39.4 to 42.1 Å (**Figure 2.11*A***). The dimer and tetramer interfaces in the tetradecamer have average surface areas of 992.3 Å^2 and 501.1 Å^2 respectively (**SI Table 2.2**), similar to the interface surface

areas seen in the tetramer structure. In addition, the hydrogen bonding and salt bridges of the interfaces are formed by the same amino acids in both structures (**Figure 2.9**).

The tetradecamer presented here is the largest oligomeric HARP assembly to be described so far. Similar to the smaller HARP dodecamers, the Hth1307 tetradecamer forms radial assemblies. Despite the additional dimer compared to the dodecamers, the Hth1307 tetradecamer maintains a roughly 42 Å distance from the active site to the proposed binding site. The next section will closely address the similarities and differences of all HARP oligomeric structures.
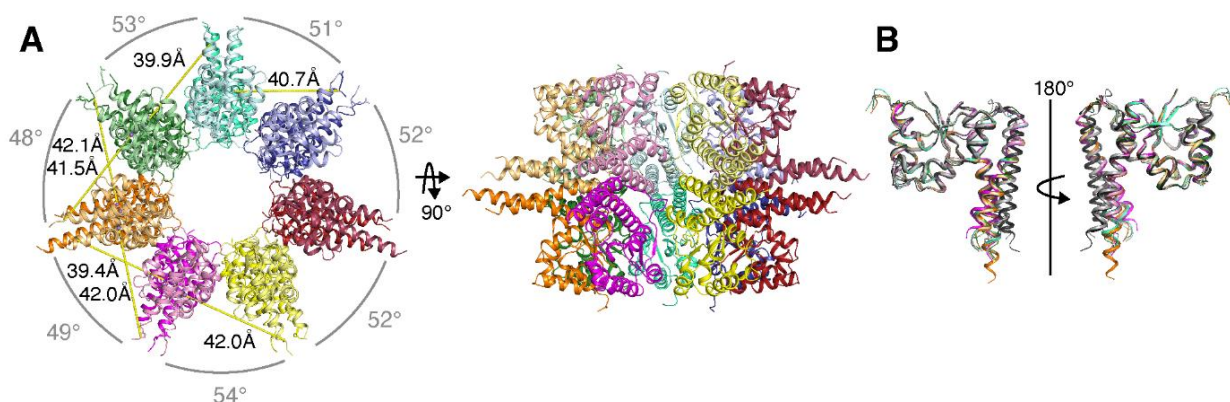
Figure 2.11: Crystal structure of the Hth1307 tetradecamer.

A. Top view of the Hth1307 14mer. Measurements from the divalent cation in the active site modeled based on alignment with *At* PRORP1 (4G24) to R125 in the putative binding site were calculated in PyMOL. Rotation of the complex by 90° along the x-axis gives a side view that shows the complex is flat with no gaps between tetramer interfaces. **B.** Alignment of unique monomers from the Hth1307 tetramer and tetradecamer structures was performed in PyMOL, with RMSD values ranging from 0.167 to 1.158 Å over 1036 to 1076 atoms. The two monomers in grey correspond to those from the Hth1307 tetramer structure, and the other seven monomers have the same coloring shown in **A**.

2.3.4 Structural comparison of available HARP structures.

Structure alignments were used to compare our structures of Hth1307 to Hhal2243 (7OG5 (19)), Aq880 (7F3E (18)), TcHARP (7E8J (20)), and PbHARP (7E8K and 7E8O (20)) using PyMOL, shown in **Figure 2.12***A*. Accompanying sequence alignments with secondary structure information can be found in **Figure 2.9**. Comparison of the residues involved in the dimer and tetramer interfaces using PISA (46) demonstrates that the interactions are predominantly

hydrophobic. Our Hth1307 structures have conformations that most closely resemble 7OG5, 7E8K, and 7E8J, which have a continuous α7 in the spike helix, rather than 7F3E and 7E8O, which have a split α7 helix due to the rearrangement of a conserved GI(I/L)DS motif that may better orient D140 for metal ion binding in the active site.

2.3.4.1 Comparison of all HARP monomers.

When comparing all available monomers (a total of 34 unique monomers, not shown) aligned by the metallonuclease domain (resi 1-73+156-189 in PyMOL), it becomes clear that the structure of the metallonuclease domain is highly conserved and the majority of the structural differences reside in the spike helix domain (**Figure 2.12***A*). The structures 7F3E and 7E8O feature a rearrangement of the highly conserved GI(I/V)DS motif in the SH domain that alters the orientation of D140 in the active site, bringing it closer to the position of D497 in *At* PRORP1. It has been proposed that this rearrangement is required for activity as it is likely that this conformational shift in HARP enzymes repositions D140 to better chelate two Mg^{2+} ions in the active site as seen in *At* PRORP1. The rearrangement of HARP would enable the binding of the equivalent of the second Mg^{2+} ion in PRORP1, which is reported to be the Mg used in catalysis. Therefore, we will refer to the structures with the rearrangement as being in the active conformation and those without to be in the inactive conformation.

For clarity, we will examine the differences between 7E8O and 7E8K (PbHARP) to investigate differences between the structures with and without the rearrangement. We observe minor differences in the beta strand at the dimer interface leading into the SH domain. The first helix of the SH domain (α6) is offset by an angle of 15.3° between the active and inactive conformations (**Figure 2.12***B*). In addition to this first helix being shifted the second helix of the SH domain in the active conformation structures is split into two smaller alpha helices (α7 & α8)

43

due to the rearrangement of the GI(I/V)DS motif (**Figure 2.12*C***). The α7 helix of the active conformation aligns with the lower half of α6 in the inactive conformation, and α8 of the active conformation aligns with the top half of α7 in the inactive conformation.
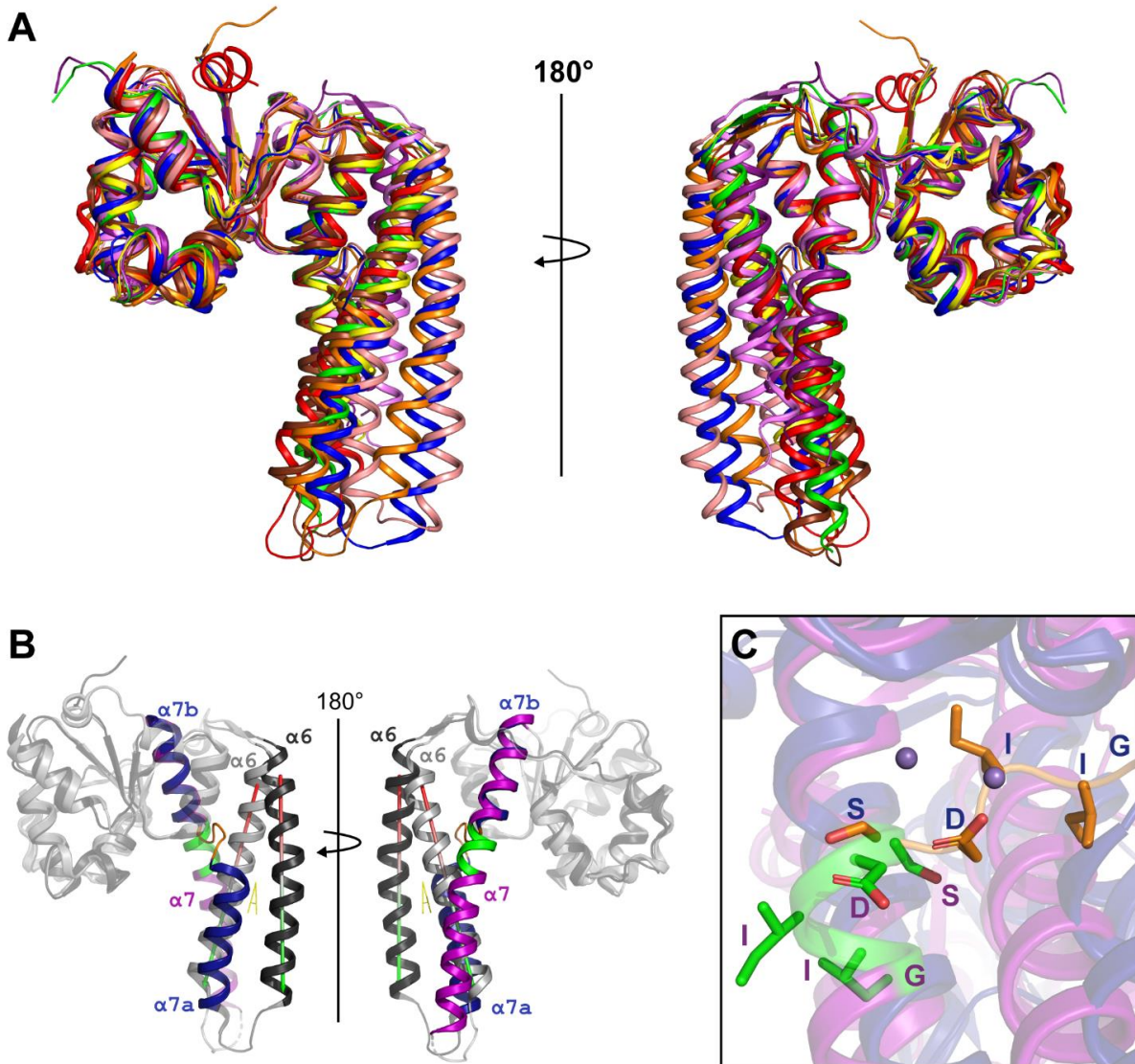
Figure 2.12: Comparison of HARP monomers.

A. Monomers in violet, deep purple, and green are Hth1307. Yellow is Hhal2243 (PDBID: 7OG5). Blue is Aq880 (PDBID: 7F3E). Brown is Tc HARP (PDBID: 7E8J). Red is Pb HARP without tRNA bound (PDBID: 7E8K). Orange and salmon are Pb HARP with tRNA bound (PDBID: 7E8O). Although all 34 unique monomers were aligned, those with significant superposition with other monomers were excluded for ease of viewing. **B**. Rearrangement of the conserved GI(I/L)DS motif in the spike helix domain in *Pb* HARP. The structure of inactive *Pb* HARP (PDBID: 7E8K) is colored in grey, purple, and green. The structure of active *Pb* HARP (PDBID: 7E8O) is colored in grey, dark blue, and orange. **C**. Zoomed in rearrangement of the same motif highlighting the GI(I/L)DS side chains. The manganese ions were modeled from the alignment of the *At* PRORP1 active site to the *Pb* HARP active site and are shown in 0.25x scale for clarity.

2.3.4.2 Comparison of all HARP dimers.

We next aligned all available dimers (a total of 19 unique dimers) by the metallonuclease domain of one chain to investigate if there is any variation in the dimer conformations, such as the relative orientation of the two protomers in the dimer or in the dimer interface. We note that the dimer interface residues determined using EMBL PISA are conserved structurally despite differences in sequence identity (**Figure 2.9** and**Figure 2.10**). The interfaces are primarily formed by interactions between the spike helices of the two monomers, with some interactions between the beta strands at the top of the SH domain (**Figure 2.8***B*).

2.3.4.3 Comparison of all HARP tetramers.

We also aligned all individual tetramers that are found in all available structures, which demonstrated the breadth of conformational flexibility within the HARP structures. The two tetramers that stand out the most are the ones that form the closing interface of 7F3E (Hhal2243) and 7OG5 (Aq880), with one of the dimers shifted ~20 Å to form a staggered interface. Tetramers with the spike helix/active site rearrangement have nearly identical dimer and tetramer interfaces to all the other structures with the exception of a salt bridge formed between Y87 and K140 in Aq880 and Y101 and K153 in Pb which is not seen in tetramers without the rearrangement (**SI Video 2.1**). The structures without the rearrangement have Y87 forming a salt bridge with D146, so this change may be required for D146 to be able to flip into the active site. Beyond these differences, the tetramer interface is nearly identical in all the structures.

2.3.4.4 Comparison of all HARP higher-order oligomers.

When comparing higher-order HARP complexes, there is a significant deviation of our structure compared to previously solved structures, as seen in **Figure 2.13**. The published models of dodecameric HARP adopt a left-handed helical structure that causes the dimers involved in the closing interface to be staggered by approximately 20 Å. Strikingly, unlike all of the dodecameric structures, our tetradecameric structure has no helical features and is instead a flat ring with nearly identical tetramer interfaces throughout. It is unclear whether the helical structures or the flat ring structure are biologically relevant or required for activity.
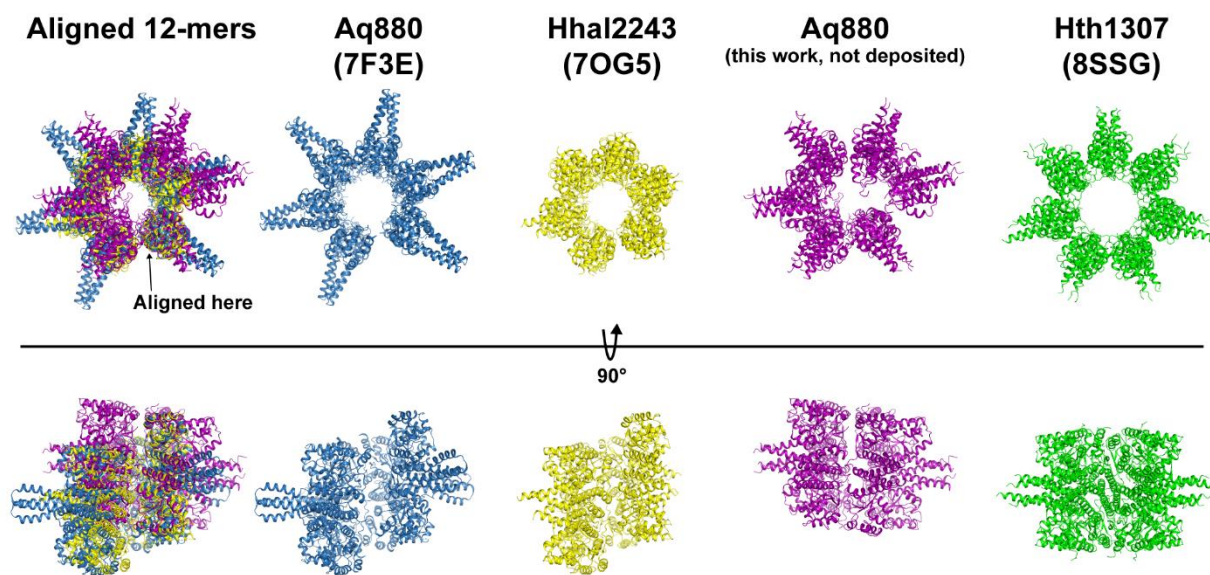
Figure 2.13: Comparison of HARP dodecamers and tetradecamer.

Alignment of the three available dodecameric structures of HARP next to the Hth1307 tetradecamer from two angles shows the differences in the helical turn and closing interface of the four structures.

2.4 Conclusions

Here, we present cryo-EM and X-ray crystallography structures of *A. aeolicus* and *H. thermophilus* HARP enzymes. The structures of HARP as a dimer, tetramer, and dodecamer have

been reported by different labs and recapitulated here (22–24). However, our crystal structure of Hth1307 is the first to reveal an even larger oligomer: a 14-mer. The overall structure of the 14-mer presented here is a heptamer of dimers, similar to the hexamer of dimers seen in the dodecameric structures previously reported . The dimer and tetramer interfaces of the 14-mer are highly similar to the interfaces seen in previously published HARP structures. The ability of Hth1307 to form a 14-mer in vitro was confirmed using native MS experiments, providing evidence that a 14-mer structure is not an artifact of our crystallization conditions. Previous publications had established that HARPs can exist as monomers, dimers, tetramers, and dodecamers, but our data indicate that HARPs have a broader range of oligomeric states than previously considered, which may be organism specific. It is unknown what contributes to these varied oligomeric states or their physiological and functional relevance in thermophiles that encode HARP. The propensity for ring-like protein structures in thermophiles and other extremophiles has been previously noted. In *Aquifex aeolicus* specifically, HARP, RNase PH (57), and L-seryl-tRNA^{Sec} selenium transferase (58) all form oligomeric ring structures. It has been proposed that proteins in thermophilic organisms may form oligomeric structures to improve thermostability (59). Because HARP requires at least a tetramer for catalysis (19), which would leave multiple interfaces open to solution, it may be more energetically favorable to form a higher order oligomer with large interface surface areas.

2.4.1 Similarities across HARP structures

Comparison of all available HARP structures reveals similar structural features such as the spike helix domain, a conserved active site, and similar dimer and tetramer interfaces. Briefly, the active sites all contain four aspartate residues and one glutamate residue, which are conserved across HARPs and poised to coordinate two magnesium ions for catalysis, although there is no

structural data to confirm this yet. All HARP structures also share the same four-helical bundle or coiled coil formed by two spike helix domains at the dimer interface. The dimer and tetramer interfaces are primarily formed by hydrophobic contacts, and while the residues have sequence and positional variability, they have conserved function.

2.4.2 Discrepancies in the oligomeric state of HARPs

At present, ten individual structural models of HARP enzymes have been published, although three data sets had poor resolution and were not made publicly available. The oligomeric states of the captured structures are varied, including one dimer, three tetramers, five dodecamers, and one tetradecamer (18–20). Interestingly, Hth1307 was characterized as a dodecamer by Teramoto et al, who used negative stain to confirm this oligomeric state (18), and as a tetramer and tetradecamer in this work, as confirmed by native mass spectrometry and x-ray crystallography. This highlights that differences in oligomerization are not due to the species of HARP or the exact amino acid sequence. It is currently unclear what causes these differences in oligomeric state, although comparing the protein purification protocols used by the different groups (**Figure 2.14**) demonstrates significant variation of the conditions used to isolate HARP enzymes from a variety of species. Because pH and salt concentrations are known to alter side chain interactions that may be involved in dimer or tetramer interface formation, it is possible that differences in these conditions lead to differences in oligomerization. For example, in the case of Hth1307, the dodecamer was purified using a lysis buffer at a pH of 8.0 with 500 mM NaCl, while the work here purified a tetradecamer using a lysis buffer at a pH of 7.8 with 1 M NaCl. However, a systematic investigation of how buffer conditions has not yet been done to determine how these conditions influence the oligomerization of HARPs.
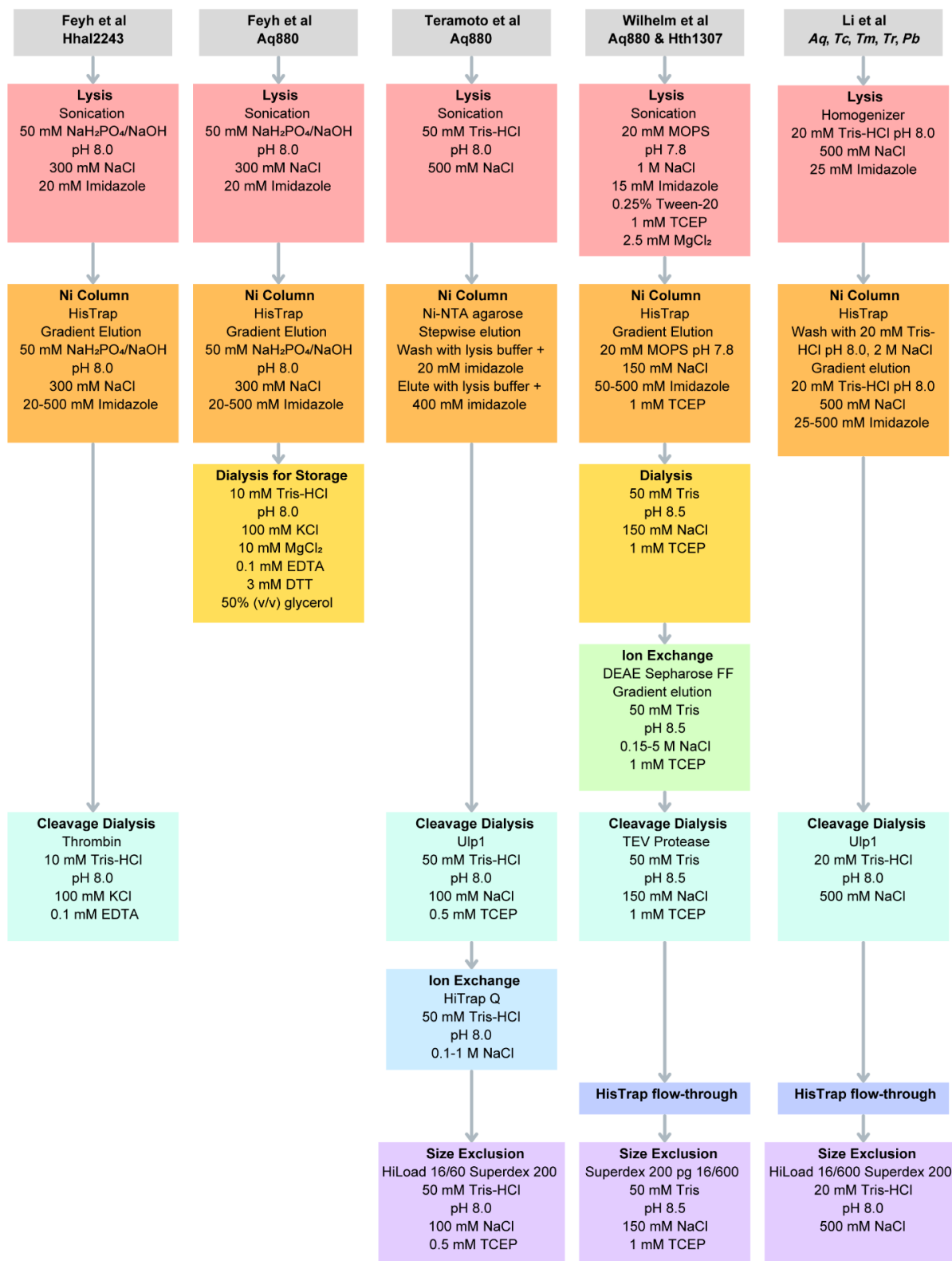
Figure 2.14: Comparison of HARP purification protocols.

Five different purification protocols were compiled into flow charts to compare purification methods and conditions used throughout protocols. The significant differences between protocols may contribute to the different oligomeric states of HARP observed in different reports.

2.4.3 Discrepancies in the structures of dodecameric and tetradecameric HARPs

Of the HARP structures published so far, only two other high-resolution structures of higher order HARP oligomers have been made publicly available. When comparing the previously published Aq880 (18) and Hhal2243 (19) cryo-EM structures to the dodecameric structure of Aq880 in this work, inconsistencies appear. For example, the Aq880 structure published by Teramoto et al (PDBID: 7F3E) and the Hhal2243 structure published by Feyh et al (PDBID: 7OG5) each adopt a left-handed screw-like formation, while the Aq880 structure presented in this work adopts a right-handed screw-like formation. These slightly helical formations are also in contrast to the structure of tetradecameric Hth1307, which forms a completely flat, symmetrical ring-like structure. Based on these data, we question whether HARP enzymes form a staggered interface and an overall helical structure under biological conditions. It is unusual for biomacromolecules with similar or the same chemical makeup to form helical structures with different handedness. The only readily available example of a molecule capable of forming both right- and left-handed helices is DNA. However, the vast majority of DNA is right-handed. Left-handed, or Z-DNA, only occurs in sequences of purine-pyrimidine repeats (such as GCGCGC) under high salt concentrations in vitro or during chromatin remodeling, supercoiling, and stabilization by Z-DNA-binding-proteins in vivo (60). The difference in the handedness of the helical HARP structures, especially with Aq880 forming helical structures with both right- and left-handedness, implies that other factors could be at play. Because the HARP structures with a helical formation were all determined using Cryo-EM, it is possible the helical formation in these structures is a result of the air-water interface problem in Cryo-EM. Simply put, it has been shown that proteins can diffuse from the bulk-phase to adsorption at the air-water interface in less than a millisecond, which can lead to orientation bias, partial or full denaturation, and dissociation of

complexes (61–65). However, it is likely that both the flat and helically distorted structures are thermodynamically accessible. It is possible that the helically distorted oligomers offer insight as to the assembly of HARP oligomers and the conformations captured represent an intermediate stage in expanding the oligomer by the addition of a dimer in the staggered interface. To learn more about the structure of HARP in solution, it may be worthwhile to use crosslinking as a tool to "lock in" the structure of HARP prior to preparing Cryo-EM grids to eliminate minimize the effects of the air-water interface problem.

2.4.4 Lack of consensus for tRNA binding

The precise orientation and interactions required for tRNA binding to HARPs remain unknown, although significant efforts have provided clues leading to the proposal of two binding models. The field agrees that a tetramer is the minimal oligomeric state required for tRNA binding, as truncations of HARP that prevent tetramer formation have been shown to have drastically reduced or zero cleavage activity. Along these lines, it is also agreed that substrate pre-tRNA likely interacts with a metallonuclease domain in one dimer and the spike helix (SH) domain of the neighboring dimer, as the distance between these domains has been shown to be ~42 Å in multiple structures and mutation of positively charged residues in the SH domain cause the complex to have reduced or zero cleavage activity. However, mutagenesis studies of the SH domain remain incomplete.

Feyh et al determined the Aq880 mutant R125A had reduced activity and R129A, R125A/R129A, and K119A/R123A/R125A/K127A/R129A had zero activity (19). Because the R129A single mutant had no activity, it is impossible to glean any information about the other residues tested in the construct with five mutations. The residues tested in these experiments are

in α7 of the SH domain, but a similar motif of four conserved, positively charged amino acids can be found in α6 of the SH domain. In Aq880, these are R94, R96, K99, and R102.

Another publication describes a co-crystal structure of tetrameric Pb HARP and *E. coli* pre-tRNA^{His} obtained by Li et al (20). In this model, the phosphate backbone of bases G52, U53, U54, C55, and G56 in the elbow region of the pre-tRNA form hydrogen bonding contacts with R116', R123', T135, R138, and R142 in the SH domains of one dimer (20). These findings were complemented by binding assays demonstrating mutation of many of these residues in the SH domain significantly decreased the binding affinity of Pb HARP (20). However, this binding model comes with several caveats. While the structure clearly shows the elbow region of the tRNA interacting with the spike helix domain of a tetramer, the acceptor stem containing the cleavage site juts out into empty space with no protein contacts. Manually modeling a copy of the tetramer with one of the dimers aligned to the original dimer interacting with the pre-tRNA provides some insight as to how the tRNA may fit in the proposed binding site (**Figure 2.15*A***). This model suggests up to 12 tRNAs could bind to dodecameric HARP. However, there is significant clashing between the tRNA and the active site and α7 of the subunit acting as the active site. In addition to these structural issues, the authors state that Pb HARP purified only as a dimer, which is unlike any of the other characterized HARPs, cleaved only 6% of *T. thermophilus* pre-tRNA^{Gly} at 20 minutes as opposed to Aq880 cleaving 96% of the substrate, and Pb HARP had a dissociation constant (K_D) of ~4 μM with *E. coli* pre-tRNA^{His} (20), while Feyh et al observed a single turnover K_M of 33 nM and a multiple turnover K_m of 12 nM (19). The extreme deviation in oligomeric state, cleavage activity, and binding affinity of Pb HARP compared to other HARPs raises questions about whether the published co-crystal model accurately represents how HARPs bind pre-tRNA

in α7 of the SH domain, but a similar motif of four conserved, positively charged amino acids can be found in α6 of the SH domain. In Aq880, these are R94, R96, K99, and R102.

Another publication describes a co-crystal structure of tetrameric Pb HARP and *E. coli* pre-tRNA^{His} obtained by Li et al (20). In this model, the phosphate backbone of bases G52, U53, U54, C55, and G56 in the elbow region of the pre-tRNA form hydrogen bonding contacts with R116', R123', T135, R138, and R142 in the SH domains of one dimer (20). These findings were complemented by binding assays demonstrating mutation of many of these residues in the SH domain significantly decreased the binding affinity of Pb HARP (20). However, this binding model comes with several caveats. While the structure clearly shows the elbow region of the tRNA interacting with the spike helix domain of a tetramer, the acceptor stem containing the cleavage site juts out into empty space with no protein contacts. Manually modeling a copy of the tetramer with one of the dimers aligned to the original dimer interacting with the pre-tRNA provides some insight as to how the tRNA may fit in the proposed binding site (**Figure 2.15*A***). This model suggests up to 12 tRNAs could bind to dodecameric HARP. However, there is significant clashing between the tRNA and the active site and α7 of the subunit acting as the active site. In addition to these structural issues, the authors state that Pb HARP purified only as a dimer, which is unlike any of the other characterized HARPs, cleaved only 6% of *T. thermophilus* pre-tRNA^{Gly} at 20 minutes as opposed to Aq880 cleaving 96% of the substrate, and Pb HARP had a dissociation constant (K_D) of ~4 μM with *E. coli* pre-tRNA^{His} (20), while Feyh et al observed a single turnover K_M of 33 nM and a multiple turnover K_m of 12 nM (19). The extreme deviation in oligomeric state, cleavage activity, and binding affinity of Pb HARP compared to other HARPs raises questions about whether the published co-crystal model accurately represents how HARPs bind pre-tRNA

substrates in solution and if this structural binding model can be applied to HARP from other organisms due to the differences in oligomerization.

To obtain a binding model, we aligned the active sites of our Hth1307 structures with the active site of the cryo-EM structure of the human mitochondrial protein-only RNase P in complex with mitochondrial pre-tRNA^{Tyr} (PDBID: 7ONU) to gain insight into how HARPs may bind pre-tRNAs (**Figure 2.15*B-C***) (50). Based on this alignment, it appears the active site of Hth1307 HARP is oriented such that the T and D loops of a pre-tRNA substrate would interact with the SH domain of the protomer diagonally opposite the active site (**Figure 2.15*D***), with a maximum of one substrate binding per tetramer interface, for a total of up to 7 tRNAs binding the 14-mer. This could explain why a tetramer is the minimum catalytic unit of HARP enzymes. We do note clashing of the pre-tRNA with HARP in our alignment, specifically between the α6 helix of one protomer and the acceptor stem of the tRNA and between the α6 helix of the diagonally opposite protomer and the T and D loops. This suggests a structural rearrangement of the SH domain is required for our proposed binding model to be correct. A structure of HARP complexed with a pre-tRNA that inserts into the active site is required to confirm either of the proposed binding models.
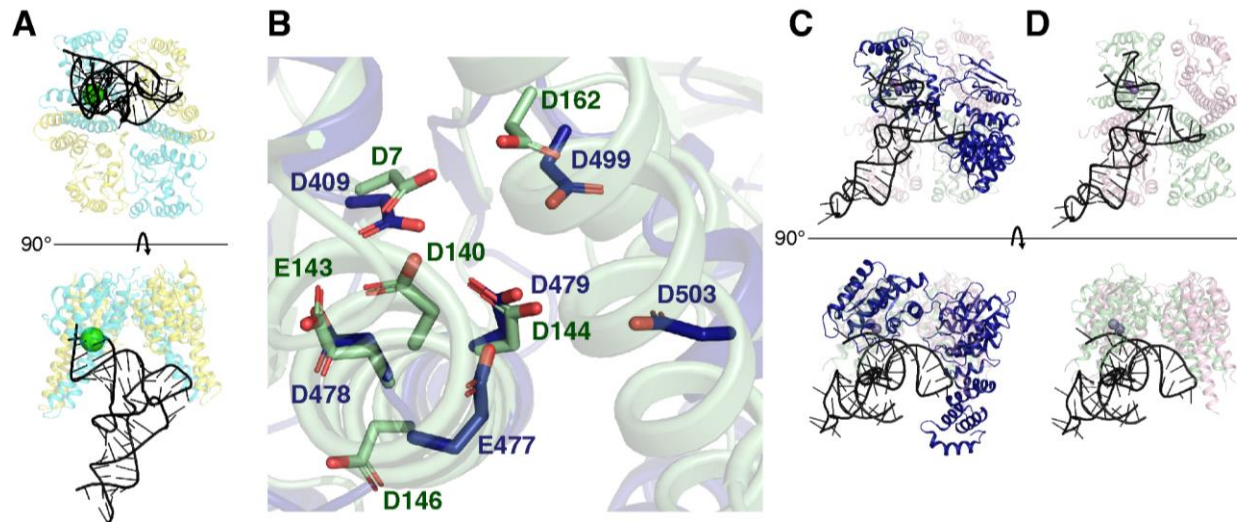
Figure 2.15: Proposed HARP binding models.

A. Pb HARP (PDBID: 7E8O) is shown in cyan and yellow and is modeled with *E. coli* pre-tRNA[His] based on the binding model presented by Li et al (2022). The calcium ion in this structure is enlarged 1.5x to showcase the active site. **B.** Active site residues of MRPP (dark blue, PDBID: 7ONU) and Hth1307 tetradecamer (pale green). **C.** Two views of the structural alignment of a tetramer from the Hth1307 tetradecamer (transparent pink and pale green) to MRPP3 complexed with mt-pre-tRNA[Tyr]. Metal ions were modeled based on manganese ions in the At PRORP1 structure (PDBID: 4G24) and are enlarged 1.5x to showcase the active site. **D.** Hth1307 modeled with mt-pre-tRNA[Tyr] spanning to interact with diagonally opposite protomers. Metal ions were modeled based on manganese ions in the At PRORP1 structure (PDBID: 4G24) and are enlarged 1.5x to showcase the active site.

2.5 Supplementary Information

SI Table 2.1: X-ray data collection and refinement statistics for Hth1307.

	Hth1307 tetramer	Hth1307 tetradecamer
Data collection		
Space group	$C222_1$	$P2_122_1$
Cell dimensions		
a, b, c (Å)	90.5, 109.2, 106.1	102.3, 113.7, 155.3
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	69.72 (2.50)	29.86 (3.20)
R_{sym} or R_{merge}	30 (396)	12 (196)
$I / \sigma I$	6.1 (1.3)	12.1 (1.2)
Completeness (%)	100 (100)	100 (99.8)
Redundancy	8.5 (8.6)	6.8 (7.2)
Refinement		
Resolution (Å)	48.45(2.60)	29.74 (3.20)
No. reflections	18585/2076	30570/1542
R_{work} / R_{free}	21.1/25.4	16.2/23.7
No. atoms		
Protein	2877	19967
Ligand/ion	15	
Water	12	-
B-factors		
Protein	76.25	116.4
Ligand/ion	99	
Water	63	
R.m.s. deviations		
Bond lengths (Å)	0.010	0.013
Bond angles (°)	2.103	1.890

SI Table 2.2: PISA interface data for both Hth1307 structures.

Interface	Chain 1	$^iN_{at}$	$^iN_{res}$	Surface, A^2	Chain 2	$^iN_{at}$	$^iN_{res}$	Surface, A^2	Interface area, A^2	Δ^iG (kcal/mol)	Δ^iG p-value	N_{HB}	N_{SB}	N_{DS}	CSS
-16: **Hth1307 14mer (8SSG)**																
Dimer	B	130	34	10244	A	127	32	9993	898.2	-12.7	0.026	10	8	0	1
	D	157	39	10100	C	148	40	10868	1074	-12.2	0.06	15	11	0	1
	F	135	33	10449	E	141	36	10017	939.6	-15	0.012	8	8	0	1
	H	154	37	10567	G	153	37	10573	1121.9	-14.2	0.023	8	12	0	1
	J	138	36	10021	I	136	33	10446	941.1	-15.1	0.011	8	8	0	1
	L	147	40	10872	K	157	39	10100	1074.1	-12.3	0.059	15	11	0	1
	N	128	32	10003	M	129	34	10241	897.1	-12.6	0.026	10	8	0	1
								Average:	992.3	-13.4	0.031	10.6	9.4	0	1
Tetramer	C	60	12	10868	A	83	24	9993	504.8	-2.7	0.558	4	0	0	0
	D	79	25	10100	B	61	13	10244	519.7	-4.9	0.327	6	3	0	0
	E	64	13	10017	C	81	22	10868	510.2	-2.9	0.463	4	0	0	0
	F	79	23	10449	D	63	13	10100	487.1	-3.9	0.407	5	0	0	0
	G	60	12	10573	E	83	23	10017	507.9	-5.5	0.271	3	2	0	0
	H	80	23	10567	F	65	12	10449	492.6	-3.4	0.46	6	0	0	0
	I	63	12	10446	G	80	24	10573	494.1	-3.4	0.452	6	0	0	0
	J	81	23	10021	H	60	12	10567	508.6	-5.4	0.271	3	2	0	0
	K	62	13	10100	I	80	23	10446	486.8	-3.8	0.41	5	0	0	0
	L	84	22	10872	J	64	13	10021	510.7	-2.9	0.469	4	0	0	0
	M	61	13	10241	K	79	25	10100	520.8	-4.9	0.328	6	3	0	0
	M	76	22	10241	A	61	11	9993	482.6	-4.9	0.375	4	0	0	0
	N	83	24	10003	L	60	12	10872	505.7	-2.7	0.559	4	0	0	0
	N	60	11	10003	B	76	22	10244	484.1	-4.9	0.366	4	0	0	0
								Average:	501.1	-4.0	0.408	4.6	0.7	0	0
-16: **Hth1307 tetramer (8SSF)**																
Dimer	B	150	36	10003	A	139	34	9981	1459.3	-27.1	0.003	5	3	0	1
Tetramer	B	61	11	10003	A	75	21	9981	589.5	-6.9	0.192	0	2	0	0.02

$^iN_{at}$	the number of interfacing atoms in the corresponding structure
$^iN_{res}$	the number of interfacing residues in the corresponding structure.
Surface, A^2	the total solvent accessible surface area in square Ångstroms
Interface area, A^2	calculated as difference in total accessible surface areas of isolated and interfacing structures divided by two.
Δ^iG (kcal/mol)	the solvation free energy gain upon formation of the interface
Δ^iG p-value	the P-value of the observed solvation free energy gain

56

N_{HB}	the number of potential hydrogen bonds across the interface
N_{SB}	the number of potential salt bridges across the interface
N_{DS}	the number of potential disulfide bonds across the interface
CSS	the Complexation Significance Score, which indicates how significant for assembly formation the interface is

SI Table 2.3: PISA interface data for all other published HARP structures.

								Interface	Δ^iG					
Aq880 12mer (7F3E)														
Interface	Chain 1	$^iN_{at}$	$^iN_{res}$	Surface, A²	Chain 2	$^iN_{at}$	$^iN_{res}$	Surface, A²	Interface area, A²	Δ^iG (kcal/mol)	Δ^iG p-value	N_{HB}	N_{SB}	N_{DS}	CSS
Dimer	A	206	47	11144	B	204	45	11152	2013.8	-31.9	0.007	14	11	0	0
	C	212	48	11141	D	212	48	11141	2099	-33.1	0.006	15	12	0	0
	E	214	49	11154	F	214	49	11155	2142.8	-33.5	0.006	15	12	0	0
	G	212	49	11154	H	216	49	11149	2132.2	-33.4	0.006	15	12	0	0
	I	214	48	11145	J	212	48	11143	2097.1	-33.1	0.006	15	12	0	0
	K	203	45	11141	L	208	47	11149	2019.3	-32	0.007	15	11	0	0
								Average:	2084.1	-32.8	0.007	15	12	0	0
Tetramer	A	81	21	11144	C	80	19	11141	722.9	-6.2	0.334	5	2	0	0
	B	78	19	11152	D	79	21	11141	697.5	-7.3	0.25	6	1	0	0
	C	81	22	11141	E	80	19	11154	727.1	-7.6	0.244	5	1	0	0
	D	82	19	11141	F	82	22	11155	712.4	-6.4	0.324	5	1	0	0
	E	83	23	11154	G	79	19	11154	733.3	-6.4	0.32	5	1	0	0
	F	80	19	11155	H	82	22	11149	723.6	-6.4	0.322	5	1	0	0
	G	82	22	11154	I	81	19	11145	722.1	-6.4	0.32	5	1	0	0
	H	81	19	11149	J	81	21	11143	722.3	-7.5	0.246	5	1	0	0
	I	79	21	11145	K	79	19	11141	689.9	-5	0.403	5	1	0	0
	J	80	19	11143	L	81	22	11149	726.1	-6.3	0.322	4	2	0	0
								Average:	717.7	-6.5	0.309	5	1	0	0
Closing	A	10	3	11144	K	8	3	11141	99	-1	0.413	1	1	0	0
	B	7	3	11152	L	11	3	11149	92.7	-0.8	0.435	1	1	0	0
								Average:	95.9	-0.9	0.424	1	1	0	0
Hhal2243 12mer (7OG5)														
Interface	Chain 1	$^iN_{at}$	$^iN_{res}$	Surface, A²	Chain 2	$^iN_{at}$	$^iN_{res}$	Surface, A²	Interface area, A²	Δ^iG (kcal/mol)	Δ^iG p-value	N_{HB}	N_{SB}	N_{DS}	CSS
Dimer	A	108	29	9051	B	120	27	9366	1153	-18.8	0.021	11	0	0	0
	C	116	28	9128	D	122	28	9115	1202.7	-19.4	0.022	12	1	0	0
	E	112	25	8739	F	113	26	8799	1159.2	-22.6	0.004	9	3	0	0
	G	118	27	8809	H	113	28	8848	1156.6	-22.6	0.004	9	0	0	0
	I	122	27	8853	J	119	30	9581	1214.1	-19.2	0.012	11	2	0	0
	K	108	27	9045	L	110	26	8908	1206.7	-18.3	0.022	13	6	0	0
								Average:	1182.1	-20.1	0.014	11	2	0	0
Tetramer	A	99	25	9051	C	99	25	9128	892.1	-5.1	0.433	6	1	0	0
	B	101	24	9366	D	99	26	9115	920.3	-3.7	0.452	9	3	0	0
	C	99	26	9128	E	97	25	8739	904.8	-5.7	0.376	8	3	0	0

Interface	Chain 1	$^iN_{at}$	$^iN_{res}$	Surface, A²	Chain 2	$^iN_{at}$	$^iN_{res}$	Surface, A²	Interface area, A²	Δ^{i}G (kcal/mol)	Δ^{i}G p-value	N_{HB}	N_{SB}	N_{DS}	CSS
	D	103	26	9115	F	97	26	8799	927.6	-3.7	0.498	13	5	0	0
	E	104	23	8739	G	107	24	8809	861.8	-5	0.441	11	4	0	0
	F	101	24	8799	H	97	23	8848	866.9	-4.8	0.416	11	7	0	0
	G	96	22	8809	I	95	22	8853	851.7	-4.3	0.394	12	3	0	0
	H	108	24	8848	J	96	24	9581	940.1	-4.8	0.369	10	5	0	0
	I	100	23	8853	K	112	25	9045	918.6	-4.1	0.447	11	3	0	0
	J	96	24	9581	L	100	25	8908	880.4	-3.6	0.462	11	4	0	0
								Average:	896.4	-4.5	0.429	10	4	0	0
Closing	A	43	15	9051	L	40	12	8908	300.7	2	0.727	4	3	0	0
	A	31	7	9051	K	26	5	9045	274.6	-0.8	0.535	3	1	0	0
	B	25	6	9366	L	32	7	8908	258.4	-2.1	0.39	2	3	0	0
								Average:	266.5	-1.4	0.462	3	2	0	0

PbHARP tetramer w/ tRNA (7E8O)

Interface	Chain 1	$^iN_{at}$	$^iN_{res}$	Surface, A²	Chain 2	$^iN_{at}$	$^iN_{res}$	Surface, A²	Interface area, A²	Δ^{i}G (kcal/mol)	Δ^{i}G p-value	N_{HB}	N_{SB}	N_{DS}	CSS
dimer	B	210	46	11906	A	208	46	12100	2163.1	-35.8	0.005	21	9	0	0.672
	D	196	43	11262	C	192	44	12004	1962	-35.1	0.004	19	2	0	0.672
								Average:	2062.6	-35.4	0.004	20	6	0	0.672
tetramer	C	104	28	12004	B	101	24	11906	992.3	-6	0.436	7	2	0	0.119
	D	83	20	11262	A	94	29	12100	885.6	-8.1	0.271	5	0	0	0.119
								Average:	938.9	-7	0.354	6	1	0	0.119
tRNA	X	31	6	11475	A	26	8	12100	270.2	-3.8	0.733	8	0	0	0.08
	X	18	6	11475	B	20	7	11906	163.2	-3.9	0.502	4	0	0	0.061
	X	18	5	11475	B	23	10	11906	156.2	-4	0.542	3	0	0	0.03

PbHARP dimer (7E8K)

Interface	Chain 1	$^iN_{at}$	$^iN_{res}$	Surface, A²	Chain 2	$^iN_{at}$	$^iN_{res}$	Surface, A²	Interface area, A²	Δ^{i}G (kcal/mol)	Δ^{i}G p-value	N_{HB}	N_{SB}	N_{DS}	CSS
tetramer	D	157	38	11594	C	166	38	11998	1609.3	-28.6	0.009	10	7	0	0.301
	B	149	38	11918	A	156	38	11942	1566.1	-28.8	0.008	9	4	0	0.301
								Average:	1587.7	-28.7	0.008	10	6	0	0.301

TcHARP dimer (7E8J)

Interface	Chain 1	$^iN_{at}$	$^iN_{res}$	Surface, A²	Chain 2	$^iN_{at}$	$^iN_{res}$	Surface, A²	Interface area, A²	Δ^{i}G (kcal/mol)	Δ^{i}G p-value	N_{HB}	N_{SB}	N_{DS}	CSS
tetramer	B	167	42	11397	A	170	43	11521	1789.9	-30.7	0.003	13	7	0	1

Video description:

Morph between inactive *Planctomycetes bacterium* (Pb) HARP (PDBID: 7E8K) and active Pb HARP (PDBID: 7E8O), showcasing structural changes as a result of the GI(I/L)DS rearrangement in α7. The Asp151 residue shown in purple is part of the GI(I/L)DS motif. The middle segment shows the change in the Tyr101 hydrogen bonding network with Asp157 and Ser61 in the inactive Pb HARP structure and Lys153 in the active Pb HARP structure, with yellow lines representing hydrogen bonds. The third segment shows changes in the arrangement of the active site.

Homologous residues in *Hydrogenobacter thermophilus* (Ht) HARP:
Pb D151 = Ht D140
Pb Y101 = Ht Y87
Pb D157 = Ht D146
Pb S61 = Ht S47
Pb K153 = Ht K142

2.6 References

1. Robertson, H. D., Altman, S., and Smith, J. D. (1972) Purification and Properties of a Specific Escherichia coli Ribonuclease which Cleaves a Tyrosine Transfer Ribonucleic Acid Precursor. *Journal of Biological Chemistry*. **247**, 5243–5251

2. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*. **35**, 849–857

3. Randau, L., Schröder, I., and Söll, D. (2008) Life without RNase P. *Nature*. **453**, 120–123

4. Randau, L. (2012) RNA processing in the minimal organism Nanoarchaeum equitans. *Genome Biol*. **13**, R63

5. Serruya, R., Orlovetskie, N., Reiner, R., Dehtiar-Zilber, Y., Wesolowski, D., Altman, S., and Jarrous, N. (2015) Human RNase P ribonucleoprotein is required for formation of initiation complexes of RNA polymerase III. *Nucleic Acids Res*. **43**, 5442–5450

6. Weber, C., Hartig, A., Hartmann, R. K., and Rossmanith, W. (2014) Playing RNase P evolution: swapping the RNA catalyst for a protein reveals functional uniformity of highly divergent enzyme forms. *PLoS Genet*. **10**, e1004506

7. Boulet, L., Karpati, G., and Shoubridge, E. A. (1992) Distribution and threshold expression of the tRNA(Lys) mutation in skeletal muscle of patients with myoclonic epilepsy and ragged-red fibers (MERRF). *Am J Hum Genet*. **51**, 1187–1200

8. Hansen, F. G., Hansen, E. B., and Atlung, T. (1985) Physical mapping and nucleotide sequence of the rnpA gene that encodes the protein component of ribonuclease P in Escherichia coli. *Gene*. **38**, 85–93

9. Walker, S. C., and Engelke, D. R. (2006) Ribonuclease P: the evolution of an ancient RNA enzyme. *Crit Rev Biochem Mol Biol*. **41**, 77–102

10. Kikovska, E., Svärd, S. G., and Kirsebom, L. A. (2007) Eukaryotic RNase P RNA mediates cleavage in the absence of protein. *Proc Natl Acad Sci U S A*. **104**, 2062–2067

11. Holzmann, J., Frank, P., Löffler, E., Bennett, K. L., Gerner, C., and Rossmanith, W. (2008) RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell*. **135**, 462–474

12. Gobert, A., Gutmann, B., Taschner, A., Gößringer, M., Holzmann, J., Hartmann, R. K., Rossmanith, W., and Giegé, P. (2010) A single Arabidopsis organellar protein has RNase P activity. *Nat Struct Mol Biol*. **17**, 740–744

13. Lechner, M., Rossmanith, W., Hartmann, R. K., Thölken, C., Gutmann, B., Giegé, P., and Gobert, A. (2015) Distribution of Ribonucleoprotein and Protein-Only RNase P in Eukarya. *Mol Biol Evol*. **32**, 3186–3193

14. Howard, M. J., Lim, W. H., Fierke, C. A., and Koutmos, M. (2012) Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5′ processing. *Proceedings of the National Academy of Sciences*. **109**, 16149–16154

15. Howard, M. J., Karasik, A., Klemm, B. P., Mei, C., Shanmuganathan, A., Fierke, C. A., and Koutmos, M. (2016) Differential substrate recognition by isozymes of plant protein-only Ribonuclease P. *RNA*. **22**, 782–792

16. Zhang, J., and Ferré-DAmaré, A. R. (2016) Trying on tRNA for Size: RNase P and the T-box Riboswitch as Molecular Rulers. *Biomolecules*. **6**, 18

17. Nickel, A. I., Wäber, N. B., Gößringer, M., Lechner, M., Linne, U., Toth, U., Rossmanith, W., and Hartmann, R. K. (2017) Minimal and RNA-free RNase P in Aquifex aeolicus. *Proceedings of the National Academy of Sciences*. **114**, 11121–11126

18. Teramoto, T., Koyasu, T., Adachi, N., Kawasaki, M., Moriya, T., Numata, T., Senda, T., and Kakuta, Y. (2021) Minimal protein-only RNase P structure reveals insights into tRNA precursor recognition and catalysis. *J Biol Chem*. **297**, 101028

19. Feyh, R., Waeber, N. B., Prinz, S., Giammarinaro, P. I., Bange, G., Hochberg, G., Hartmann, R. K., and Altegoer, F. (2021) Structure and mechanistic features of the prokaryotic minimal RNase P. *eLife*. **10**, e70160

20. Li, Y., Su, S., Gao, Y., Lu, G., Liu, H., Chen, X., Shao, Z., Zhang, Y., Shao, Q., Zhao, X., Yang, J., Cao, C., Lin, J., Ma, J., and Gan, J. (2022) Crystal structures and insights into precursor tRNA 5'-end processing by prokaryotic minimal protein-only RNase P. *Nat Commun*. **13**, 2290

21. Schwarz, T. S., Wäber, N. B., Feyh, R., Weidenbach, K., Schmitz, R. A., Marchfelder, A., and Hartmann, R. K. (2019) Homologs of aquifex aeolicus protein-only RNase P are not the major RNase P activities in the archaea haloferax volcanii and methanosarcina mazei. *IUBMB Life*. **71**, 1109–1116

22. Stols, L., Gu, M., Dieckman, L., Raffen, R., Collart, F. R., and Donnelly, M. I. (2002) A New Vector for High-Throughput, Ligation-Independent Cloning Encoding a Tobacco Etch Virus Protease Cleavage Site. *Protein Expression and Purification*. **25**, 8–15

23. Marty, M. T., Baldwin, A. J., Marklund, E. G., Hochberg, G. K. A., Benesch, J. L. P., and Robinson, C. V. (2015) Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Anal. Chem.* **87**, 4370–4376

24. Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K., and Evans, G. (2018) DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D Struct Biol*. **74**, 85–97

25. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J Appl Crystallogr*. **40**, 658–674

26. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*. **66**, 486–501

27. Bricogne, G., Blanc, E., Brandi, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O. S., Vonrhein, C., and Womack, T. O. (2017) BUSTER version 2.10.4. Cambridge, United Kingodm: Global Phasing Ltd.

28. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr*. **67**, 355–367

29. Joosten, R. P., Long, F., Murshudov, G. N., and Perrakis, A. (2014) The PDB_REDO server for macromolecular structure model optimization. *IUCrJ*. **1**, 213–220

30. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res*. **35**, W375-383

31. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. (2015)

32. Punjani, A., Rubinstein, J. L., Fleet, D. J., and Brubaker, M. A. (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods*. **14**, 290–296

33. Zheng, S. Q., Palovcak, E., Armache, J.-P., Verba, K. A., Cheng, Y., and Agard, D. A. (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods*. **14**, 331–332

34. Rubinstein, J. L., and Brubaker, M. A. (2015) Alignment of cryo-EM movies of individual particles by optimization of image translations. *Journal of Structural Biology*. **192**, 188–195

35. Rohou, A., and Grigorieff, N. (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*. **192**, 216–221

36. Stagg, S. M., Noble, A. J., Spilman, M., and Chapman, M. S. (2014) ResLog plots as an empirical metric of the quality of cryo-EM reconstructions. *Journal of Structural Biology*. **185**, 418–426

37. Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., and Ferrin, T. E. (2018) UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci*. **27**, 14–25

38. Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., and Ferrin, T. E. (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci*. **30**, 70–82

39. Ruotolo, B. T., Giles, K., Campuzano, I., Sandercock, A. M., Bateman, R. H., and Robinson, C. V. (2005) Evidence for Macromolecular Protein Rings in the Absence of Bulk Water. *Science*. **310**, 1658–1661

40. Farrugia, M. A., Han, L., Zhong, Y., Boer, J. L., Ruotolo, B. T., and Hausinger, R. P. (2013) Analysis of a Soluble (UreD:UreF:UreG)2 Accessory Protein Complex and Its Interactions with Klebsiella aerogenes Urease by Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **24**, 1328–1337

41. Konermann, L., and Douglas, D. J. (1997) Acid-Induced Unfolding of Cytochrome c at Different Methanol Concentrations: Electrospray Ionization Mass Spectrometry Specifically Monitors Changes in the Tertiary Structure. *Biochemistry*. **36**, 12296–12302

42. Kaltashov, I. A., and Mohimen, A. (2005) Estimates of Protein Surface Areas in Solution by Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **77**, 5370–5379

43. Going, C. C., and Williams, E. R. (2015) Supercharging with *m* -Nitrobenzyl Alcohol and Propylene Carbonate: Forming Highly Charged Ions with Extended, Near-Linear Conformations. *Anal. Chem.* **87**, 3973–3980

44. Howard, M. J., Lim, W. H., Fierke, C. A., and Koutmos, M. (2012) Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5′ processing. *Proceedings of the National Academy of Sciences*. **109**, 16149–16154

45. Howard, M. J., Klemm, B. P., and Fierke, C. A. (2015) Mechanistic Studies Reveal Similar Catalytic Strategies for Phosphodiester Bond Hydrolysis by Protein-only and RNA-dependent Ribonuclease P. *J Biol Chem*. **290**, 13454–13464

46. Krissinel, E., and Henrick, K. (2007) Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology*. **372**, 774–797

47. Teramoto, T., Kaitany, K. J., Kakuta, Y., Kimura, M., Fierke, C. A., and Hall, T. M. T. (2020) Pentatricopeptide repeats of protein-only RNase P use a distinct mode to recognize conserved bases and structural elements of pre-tRNA. *Nucleic Acids Res*. **48**, 11815–11826

48. Imai, T., Nakamura, T., Maeda, T., Nakayama, K., Gao, X., Nakashima, T., Kakuta, Y., and Kimura, M. (2014) Pentatricopeptide repeat motifs in the processing enzyme PRORP1 in

Arabidopsis thaliana play a crucial role in recognition of nucleotide bases at TψC loop in precursor tRNAs. *Biochemical and Biophysical Research Communications*. **450**, 1541–1546

49. Gobert, A., Pinker, F., Fuchsbauer, O., Gutmann, B., Boutin, R., Roblin, P., Sauter, C., and Giegé, P. (2013) Structural insights into protein-only RNase P complexed with tRNA. *Nat Commun*. **4**, 1353

50. Bhatta, A., Dienemann, C., Cramer, P., and Hillen, H. S. (2021) Structural basis of RNA processing by human mitochondrial RNase P. *Nat Struct Mol Biol*. **28**, 713–723

51. Wu, J., Niu, S., Tan, M., Huang, C., Li, M., Song, Y., Wang, Q., Chen, J., Shi, S., Lan, P., and Lei, M. (2018) Cryo-EM Structure of the Human Ribonuclease P Holoenzyme. *Cell*. **175**, 1393-1404.e11

52. Reiter, N. J., Osterman, A., Torres-Larios, A., Swinger, K. K., Pan, T., and Mondragón, A. (2010) Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature*. **468**, 784–789

53. Carrara, G., Calandra, P., Fruscoloni, P., Doria, M., and Tocchini-Valentini, G. P. (1989) Site selection by Xenopus laevis RNAase P. *Cell*. **58**, 37–45

54. Svärd, S. G., and Kirsebom, L. A. (1993) Determinants of Escherichia coli RNase P cleavage site selection: a detailed in vitro and in vivo analysis. *Nucleic Acids Res*. **21**, 427–434

55. Lan, P., Tan, M., Zhang, Y., Niu, S., Chen, J., Shi, S., Qiu, S., Wang, X., Peng, X., Cai, G., Cheng, H., Wu, J., Li, G., and Lei, M. (2018) Structural insight into precursor tRNA processing by yeast ribonuclease P. *Science*. **362**, eaat6678

56. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Research*. **32**, W665–W667

57. Ishii, R., Nureki, O., and Yokoyama, S. (2003) Crystal Structure of the tRNA Processing Enzyme RNase PH from Aquifex aeolicus *. *Journal of Biological Chemistry*. **278**, 32397–32404

58. Itoh, Y., Bröcker, M. J., Sekine, S., Hammond, G., Suetsugu, S., Söll, D., and Yokoyama, S. (2013) Decameric SelA•tRNASec Ring Structure Reveals Mechanism of Bacterial Selenocysteine Formation. *Science*. **340**, 75–78

59. Pucci, F., and Rooman, M. (2017) Physical and molecular bases of protein thermal stability and cold adaptation. *Current Opinion in Structural Biology*. **42**, 117–128

60. Ravichandran, S., Subramani, V. K., and Kim, K. K. (2019) Z-DNA in the genome: from structure to disease. *Biophys Rev*. **11**, 383–387

61. Glaeser, R. M. (2018) Proteins, interfaces, and cryo-EM grids. *Current Opinion in Colloid & Interface Science*. **34**, 1–8

62. D'Imprima, E., Floris, D., Joppe, M., Sánchez, R., Grininger, M., and Kühlbrandt, W. (2019) Protein denaturation at the air-water interface and how to prevent it. *eLife*. **8**, e42747

63. Klebl, D. P., Gravett, M. S. C., Kontziampasis, D., Wright, D. J., Bon, R. S., Monteiro, D. C. F., Trebbin, M., Sobott, F., White, H. D., Darrow, M. C., Thompson, R. F., and Muench, S. P. (2020) Need for Speed: Examining Protein Behavior during CryoEM Grid Preparation at Different Timescales. *Structure*. **28**, 1238-1248.e4

64. Galkin, V. E., Orlova, A., Vos, M. R., Schröder, G. F., and Egelman, E. H. (2015) Near-Atomic Resolution for One State of F-Actin. *Structure*. **23**, 173–182

65. Liu, N., and Wang, H.-W. (2023) Better Cryo-EM Specimen Preparation: How to Deal with the Air-Water Interface? *J Mol Biol*. **435**, 167926

Chapter 3 Kinetic Characterization of HARPs

3.1 Introduction

Ribonuclease P enzymes are essential endonucleases responsible for cleavage of the 5' end of pre-tRNAs. Although RNase P is most commonly a ribonucleoprotein complex that uses a ribozyme to coordinate catalytically active magnesium ions, protein-only RNase Ps (PRORPs) were discovered in eukaryotes such as protists, algae, land plants, and metazoans (1–3). These PRORPs have a conserved overall molecular architecture, with positively charged pentatricopeptide repeat (PPR) region positioned ~42 Å away from a metallonuclease domain that uses aspartate residues to coordinate magnesium ions for catalysis (4, 5). The activity and substrate recognition strategy of *Arabidopsis thaliana* PRORPs (*At* PRORPs) and *Homo sapiens* mitochondrial RNase P (*Hs* (mt)RNase P) have been well characterized. Briefly, hydrogen bonding networks and π-stacking of substrate nucleotides and aromatic amino acids bind and stabilize the elbow region of the pre-tRNA substrate to position the phosphodiester bond between the -1 and +1 nucleotides at the active site of the enzyme (4–6). *A. thaliana* PRORPs have been

shown to have higher cleavage activity towards *A. thaliana* pre-tRNAs compared to mitochondrial pre-tRNAs, suggesting some difference in substrate recognition between *At* PRORPs and *Hs* (mt)RNase P despite their homology (7). *A. thaliana* PRORPs are also able to cleave tRNA-like structural elements at the 5' and 3' ends of plant mitochondrial mRNA and in plant viral genomes, indicating some promiscuity (2, 8, 9).

A new class of PRORPs was recently identified in prokaryotic archaea and bacteria, termed homologs of *Aquifex* RNase P (HARPs) (10). Many of the organisms encoding HARPs are thermophiles or halophiles. These novel PRORPs are significantly smaller than previously characterized PRORPs. Sequence alignments indicated HARPs have a similar metallonuclease domain compared to At PRORPs and Hs (mt)RNase P but lack the PPR domain used for substrate recognition. Structural investigation of HARPs have revealed monomers consist of a metallonuclease domain with two protruding α-helices termed the spike helix domain (11–13). HARPs form dimers with a buried interface between the spike helix domains. These dimers arrange into a larger radial oligomeric complex with the spike helix domains pointing outwards. The spike helix domains are proposed to play a role in substrate recognition, as discussed in **Chapter 2**. From the available structural models, it has been proposed HARPs use the same catalytic mechanism as *At* PRORPs and *Hs* (mt)RNase P, although this has not been proven. There is little information available about the substrate specificity and scope of HARPs, and the behavior of thermophilic tRNAs has yet to be investigated and compared to well characterized tRNAs such as yeast tRNA^{Phe}.

To address some of these gaps, we investigated the substrate preference of the *Hydrogenobacter thermophilus* HARP (Hth1307) protein that we structurally characterized by assessing its activity for native and non-native pre-tRNA substrates. The oligomerization of HARP

creates potential for allosteric communication between the multiple putative substrate binding sites. Therefore, we used single-turnover conditions to try to limit the number of pre-tRNAs associated with each Hth1307 complex. Single-turnover conditions ensure that only a single active site per HARP complex on average will be occupied, therefore minimizing the effects of potential allosteric communication between monomer active sites on reaction kinetics. As with other studies of PRORPs (7, 10–18), gel-based cleavage assays using 5'-fluorescently labeled pre-tRNAs were used to monitor the percentage of cleaved tRNA over time.

Upon discovering a substrate preference in Hth1307, we began investigating the secondary structure features of tRNAs from two organisms encoding HARP, one thermophile lacking HARP, and three commonly used model organisms. We next began characterizing the UV melting profile of three tRNAs to determine if the melting temperature (T_M) of tRNAs from thermophilic organisms are higher than those from non-thermophiles. Although the T_M values of *H. thermophilus* pre-tRNA^{Lys} and *Thermus thermophilus* pre-tRNA^{Gly} were significantly higher than that of *Saccharomyces cerevisiae* pre-tRNA^{Phe}, the T_M values of the two thermophilic tRNAs are at or near the optimal growing temperature of their respective organisms. This suggests the tRNAs could be adopting alternative conformations under physiological conditions, and further work is needed to understand how conditions such as temperature and salinity impact the structure and activity of tRNAs and HARPs.

3.2 Methods

3.2.1 In vitro transcription and 5' end labeling of pre-tRNAs.

Pre-tRNA DNA oligos were commercially synthesized by Integrated DNA Technologies and used as a template for in vitro transcription reaction. 5'-fluorescent labeling was performed as previously described (5, 15, 17). Reactions were carried out in 30 mM Tris-HCl pH 8.0, 2 mM spermidine, 0.01% Triton-X, 24 mM $MgCl_2$, 5 mM DTT, 4 mM ATP, 4 mM CTP, 4 mM UTP, 0.8 mM GTP, 5 mM guanosine-5'-O-monophosphorothioate (GMPS), 0.08 U/mL Superase-In, 0.12 U/mL YIPP, 0.28 mM purified T7 RNA polymerase, 0.08 μM DNA template, and 0.08 μM T7 primer. After stopping the reaction by addition of EDTA pH 8.0 and NaCl to final concentrations of 50 mM and 500 mM respectively, the samples were washed with degassed TE buffer pH 7.2 three times using an Amicon Ultra-4 10 kDa MWCO spin column. The washed pre-tRNA (~200 μl) was incubated with 20 μl of 45 mM 5-iodoacetamidofluorescein overnight at 37 °C to label the 5′ end. The reaction was stopped by addition of an equal volume of 2x Loading Dye (0.05% Xylene Cyanol Dye, 7 M urea and 0.1 M EDTA pH 8.0) and run on a 12% urea-polyacrylamide gel. The pre-tRNA was eluted into Crush-Soak Buffer (TE buffer pH 8.0, 0.1% SDS, and 0.5 M NaCl) overnight at 4 °C. The next day, the mixture was filtered, then concentrated and washed with degassed TE buffer pH 8.0 using an Amicon Ultra-15 10 kDa MWCO spin column. The pre-tRNA was ethanol precipitated, and the resulting pellet was resuspended in TE buffer pH 8.0. Unlabeled pre-tRNAs were made similarly to labeled ones, except the transcription reaction contained 4 mM GTP and excluded GMPS, in addition to omitting the 5′ end labeling step with fluorescein. The concentrations of total and labeled pre-tRNA were measured from absorbance using a Nanodrop spectrophotometer. The concentrations of total and labeled pre-

tRNA were measured with absorbance using a Nanodrop (Thermo Scientific) spectrophotometer and calculated using the extinction coefficients calculated by IDT OligoAnalyzer Tool.

3.2.2 Single-turnover cleavage assays.

Fluorescently labeled pre-tRNA substrates were refolded immediately before each 5' pre-tRNA cleavage reaction. First, pre-tRNAs were heated at 95 °C for 3 minutes and cooled down to 37°C for 15 minutes. This was followed by the addition of Cleavage Reaction Buffer (30 mM MOPS pH 7.8, 150 mM NaCl, varied $MgCl_2$, 1 mM TCEP) and incubation for 15 minutes at 37°C.

All single-turnover reactions were performed in the Cleavage Reaction Buffer at 37 °C. To initiate a reaction, a mixture of 1000 nM HARP enzyme in Cleavage Reaction Buffer was added to the 2 nM folded, fluorescently labeled pre-tRNA. For gel-based assays, reaction time points were quenched with equal volumes of 2x Loading Dye and resolved on 20% denaturing urea-polyacrylamide gels. The gels were scanned with an Amersham Typhoon imager, and results were quantified using ImageQuant TL version 10.1. Data from at least three independent experiments were analyzed using Prism 9. **Equation 1** was used to fit the data to calculate the observed rate constant (*k*) and the standard error from the fitting where *A* is the endpoint, *B* is the amplitude, and *t* is the time.

$$Fraction\ cleaved = A - Be^{-k_{obs}t} \qquad (Equation\ 1)$$

3.2.3 UV melting experiments

Pre-tRNAs were folded as described above before each experiment using a "UV melting buffer" with final concentrations of 20 mM sodium cacodylate (NaCac), 150 mM sodium chloride, and 1 mM magnesium chloride. Melting curves were obtained by heating samples at 0.06 °C/s using a Thermo Fisher Evolution™ spectrophotometer. Cooling from 100 °C to 20 °C resulted in a similar but less noisy curve compared to that obtained by heating. Data was plotted using Prism

9 using 4 neighbors on each side and 4th order smoothing to obtain 1st and 2nd derivative plots. Melting temperatures (T$_M$) were obtained by determining x-intercepts of the 2nd derivative plots from cooling.

3.3 Selection of precursor tRNA substrates.

We were then interested in using single-turnover assays to discover whether Hth1307 showed any specificity for native vs non-native pre-tRNAs or had a preference for leader sequence length. Four substrates were selected, shown in **Figure 3.1**. Briefly, *B. subtilis* pre-tRNA^{Asp} with a 5 nt leader sequence and 3' CCA (Bs ptRD$^{5\ CCA}$) has been used extensively in previous characterizations of PRORP enzymes (6, 14, 16, 19, 20), *T. thermophilus* pre-tRNA^{Gly} with a 14 nt leader sequence and no 3' trailer (Tt ptRG^{14}) has been used by other groups investigating HARPs (11), *S. cerevisiae* pre-tRNA^{Phe} with a 9 nt leader sequence and no 3' trailer (Sc ptRF^{9}) is a very stable eukaryotic tRNA commonly used in structural and enzymological studies (6, 21), and Ht ptRK is a native substrate from *H. thermophilus* with one of the most prevalent codons in its genome (22). All four of these pre-tRNAs have a 5-nt variable loop and the same number of nucleotides in the acceptor stem, anticodon stem loop, and T and D loops to limit variability in the general size or folding of the substrates. To investigate whether Hth1307 has a preferred leader sequence length for catalytic activity, we used the native substrate Ht ptRK with 5, 11, 15, and 22 nucleotide leader sequences. A 3' CCA tail was added to Ht ptRK and Bs ptRD with 5 nucleotide leader sequences to determine if the presence of the 3' CCA tail impacts the activity of Hth1307. As a control to account for the nucleotides at the cleavage site, the Sc ptRF^{9} substrate used has the same -1 and +1 nucleotides as the native substrate. The purity of the *in vitro* transcribed fluorescently labeled pre-tRNAs was confirmed using denaturing urea-PAGE, shown in **Figure 3.2**.

3.4 Oligomerization is required for Hth1307 activity.

Because structural data indicates that oligomerization is likely required for pre-tRNA binding, we were first interested in determining whether all of the oligomeric states of Hth1307 isolated via size exclusion chromatography were catalytically active using three distinct fractions: a monomer, a 14-mer, and a higher-order oligomer. This data shows that the Hth1307 monomeric fraction has no cleavage activity after two hours compared to the robust activity seen in the Aq880 dodecamer sample and the Hth1307 14-mer and higher-order oligomer peaks (**Figure 3.3**). These results confirm that a higher-order HARP assembly is required for cleavage activity.
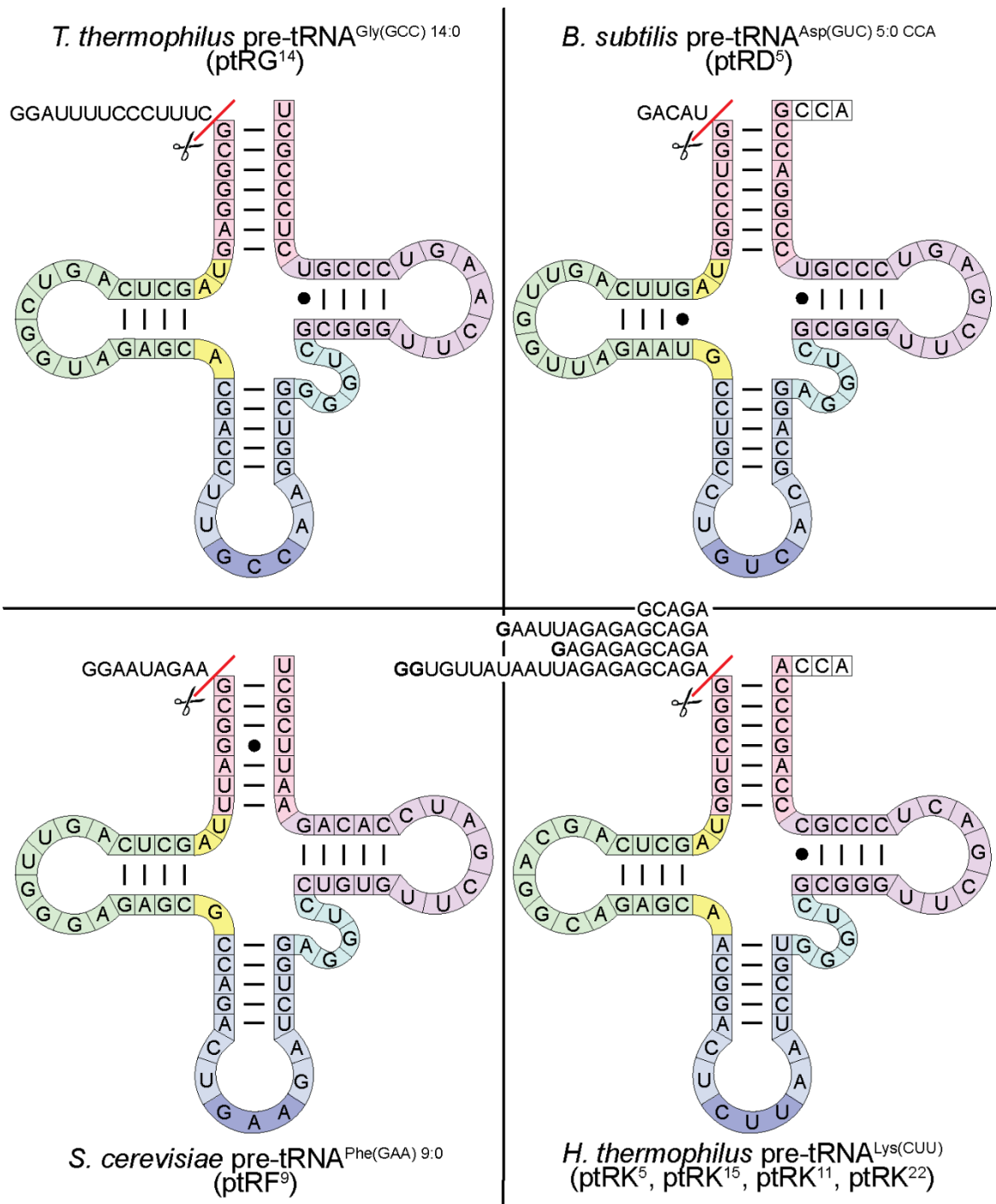
Figure 3.1: Illustrated sequences of the four pre-tRNA substrates selected for the single turnover assays.
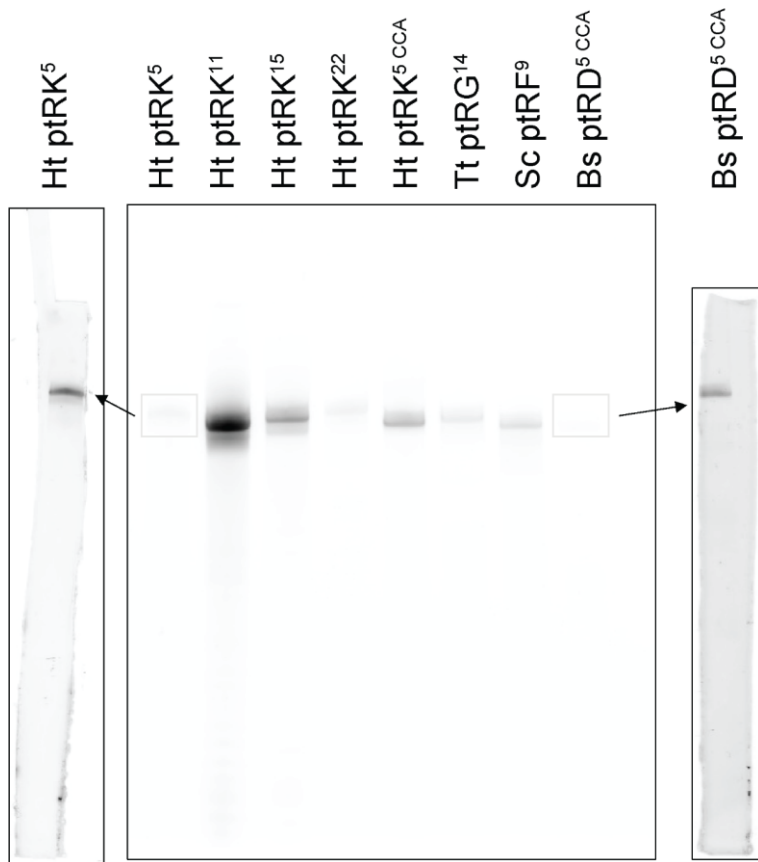
Figure 3.2: Urea-PAGE to assess substrate tRNA quality.

The quality of the pre-tRNAs used in our kinetic assays was surveyed using denaturing urea-PAGE. Due to the low labeling efficiency of Ht ptRK^5 and Bs ptRD$^{5\ CCA}$, boxed in light grey, those lanes were excised from the gel and imaged with higher amplification of the signal.
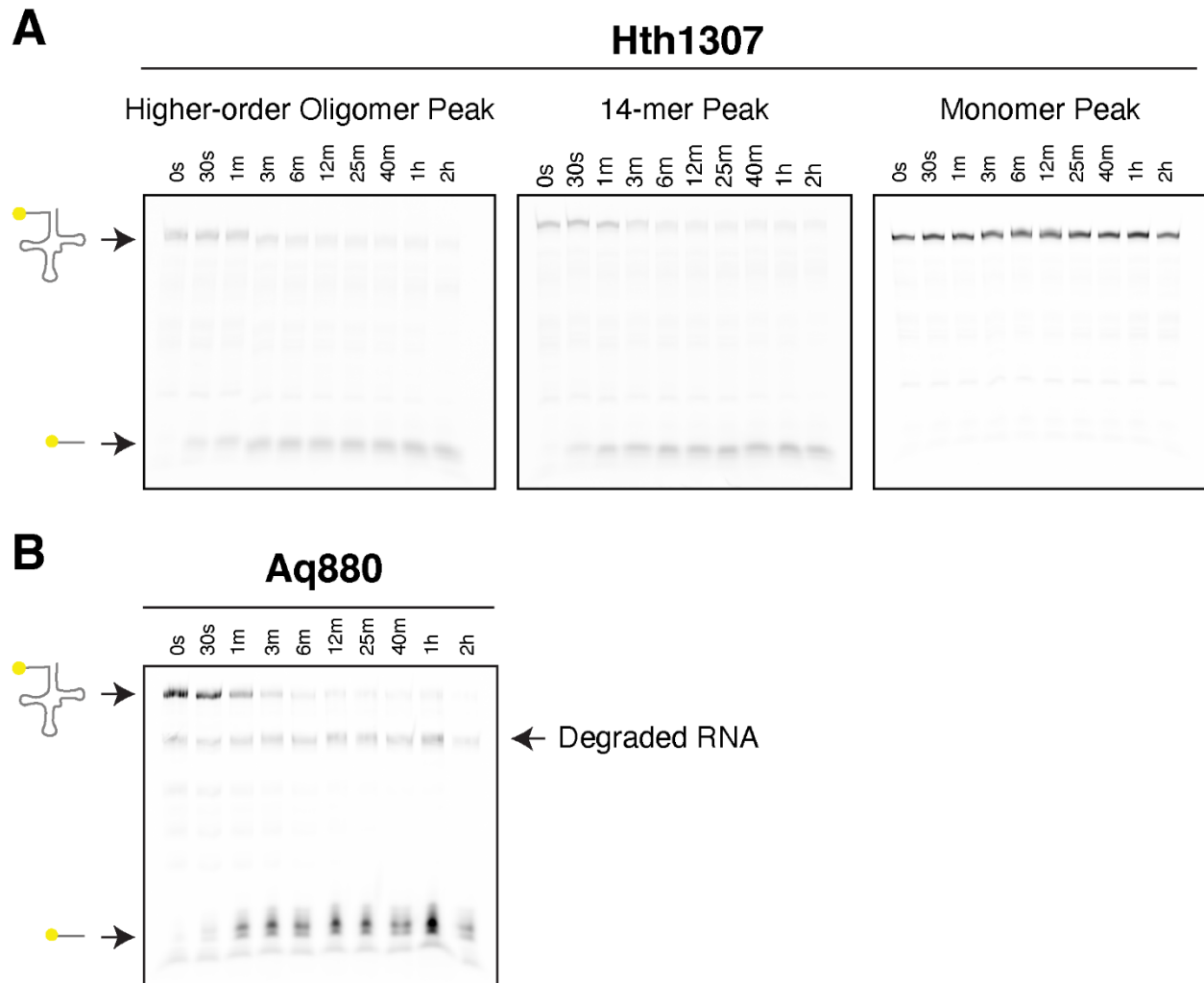
Figure 3.3: Initial confirmation of HARP activity.

Enzymatic activity of purified HARP proteins was confirmed via gel-based multiple turnover assays using **A.** 50 nM Hth1307 from different SEC peaks or **B.** 50 nM Aq880 and 5 μM Bs ptRD^5 at 5 mM $MgCl_2$ and 300 mM NaCl. The Hth1307 monomer has no activity compared to the higher order oligomers.

3.5 Selection of magnesium concentration for single turnover assays.

In eukaryotic PRORPs, it has been established that increasing the concentration of magnesium within a range may increase the rate of the reaction but also inhibit reactivity above a certain threshold (23). Therefore, it is important to establish the Mg^{2+} dependence in HARP activity because if the nuclease reaction proceeds to completion too rapidly, it is impossible to reliably calculate the observed rate constant. To investigate the effect of the catalytically required Mg^{2+} on

nuclease activity, we performed single-turnover experiments to monitor changes in activity upon varying Mg^{2+} concentrations. Using 20 nM *H. thermophilus* pre-tRNA^{Lys} with a 5-nt leader sequence (Ht ptRK[5], **Figure 3.1**) and 15,000 nM Hth1307 to ensure saturation, reactions were carried out across a range of $MgCl_2$ concentrations used to study other PRORP enzymes: 1, 2, 3, and 4.5 mM $MgCl_2$ (**Figure 3.4**) (5, 7, 10, 11, 15, 16, 18, 24). The reactions using 1 mM $MgCl_2$ enabled the measurement of time points that capture the entirety of the reaction without using stopped-flow methods. Therefore, we moved forward using 1 mM $MgCl_2$ in all subsequent assays.
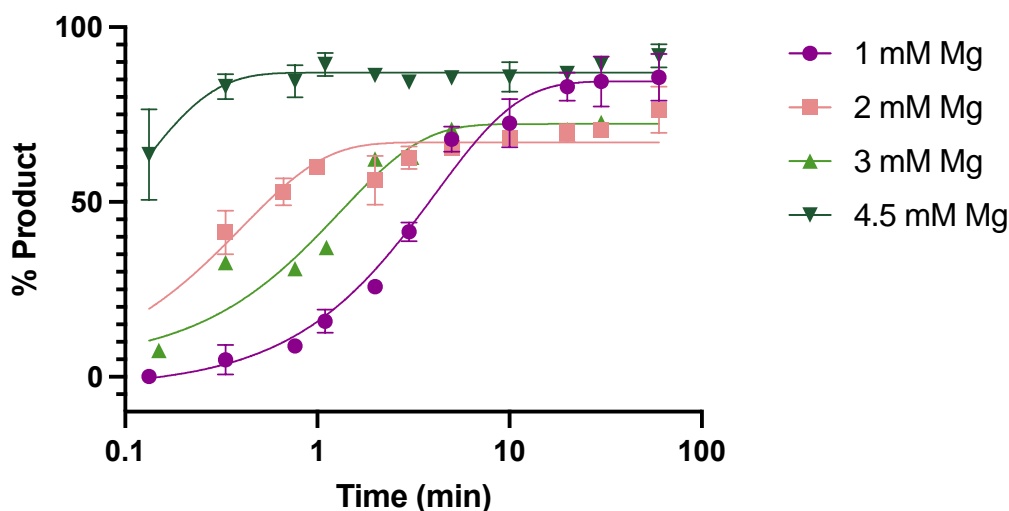
Figure 3.4: Magnesium dependence of Hth1307.

The formation of product over time at different magnesium concentrations was monitored using 20.75 μM Hth1307 and 20 nM fluorescently labeled Sc ptRF[9]. The time course for the 3 mM $MgCl_2$ reaction has only one data set, while the rest are duplicates. Data were fit to Equation 1. Error bars represent standard deviation.

3.6 Hth1307 prefers native substrates compared to non-native substrates.

We found that the activity of Ht ptRK did not vary significantly (< 2-fold) with differing 5' leader lengths (5-22 nt) (**Figure 3.5**). However, the presence of a 3' CCA sequence appears to be an important determinant of substrate selection, as the addition of this sequence slows pre-tRNA cleavage by 10-fold ($k_{obs,HTpRK5CAA} = 0.05$ min^{-1} vs $k_{obs,HTpRK5} = 0.58$ min^{-1}). In general, the enzyme inefficiently cleaved non-native pre-tRNAs, exhibiting k_{obs} values ~60-fold lower than that for

native substrates (0.01 to 0.018 min^{-1}). These results demonstrate that Hth1307 exhibits a marked preference for the native substrate.
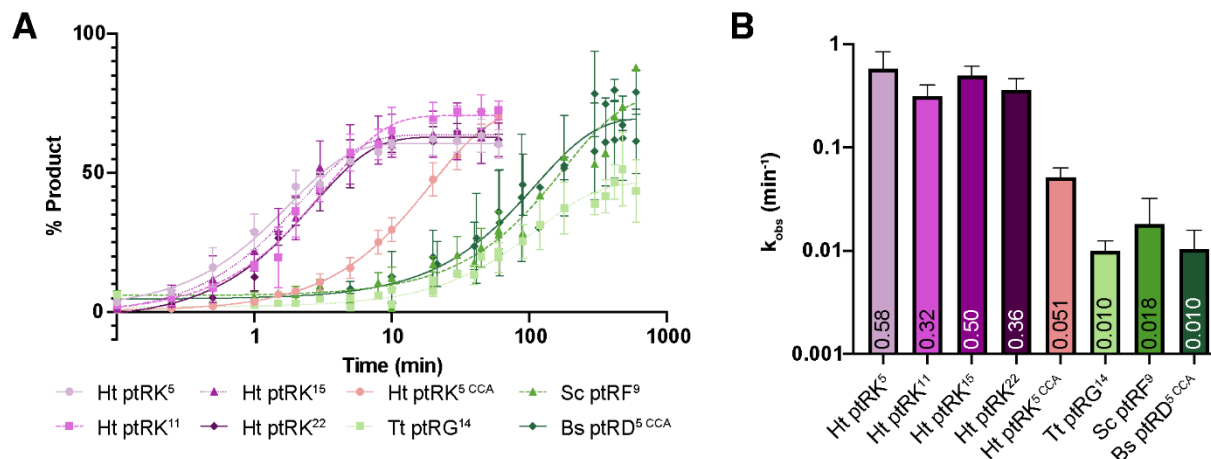
Figure 3.5: Single turnover kinetics assays using Hth1307 and various substrates.

A. Timecourse data of percent product vs time for single turnover reactions carried out with 1000 nM HARP and 2 nM of different pre-tRNA substrates. Data from at least three replicates were fit to **Equation 1**. Error bars represent standard deviation. **B.** Comparison of the observed rate constant values for the different substrates tested reveals that Ht HARP has a higher turnover rate with native substrate compared to other pre-tRNAs commonly used to study RNase P enzymes. Error bars represent standard deviation.

3.7 Examining tRNA secondary structures from different species

Unlike Hth1307, which displayed increased cleavage activity toward a specific substrate, other PRORPs have been shown to be quite promiscuous, with At PRORP having been shown to cleave even tRNA-like structures from viral genomes (2, 8, 9). To learn more about potential causes for the substrate preference seen in Hth1307, a table containing the sequence, secondary structure, and thermodynamic information was compiled to compare the features of tRNAs from six different species, found in **0**. *Aquifex aeolicus* and *Hydrogenobacter thermophilus* were included as they both contain HARPs that were used in this work. *Thermus thermophilus* was selected as an additional thermophile. Because they are well-characterized, *Escherichia coli* and *Bacillus subtilis* were selected as representative prokaryotes and *Saccharomyces cerevisiae* was selected as a representative eukaryote. By comparing the structural features of these tRNAs, we

hoped to uncover some kind of structural element that may contribute to the specificity of Hth1307 or determine if tRNAs from thermophiles have some physical property that may help explain why some organisms use HARPs in place of canonical PRORPs or the ribozyme RNase P.

The table presented in **0** shows the number of nucleotides in the acceptor stem, D stem, D loop, anticodon stem, anticodon loop, variable loop, T stem, and T loop, as well as the calculated free energy of the most optimal predicted secondary structure as determined by RNAfold (25). Sorting this data from low to high free energy, where a lower or more negative value is indicative of greater stability, there is an unsurprising trend that tRNAs from the three thermophilic organisms (*A. aeolicus*, *T. thermophilus*, and *H. thermophilus*) have a larger proportion of tRNAs with ΔG values lower than the average, likely due to a higher GC content that provides stability at the high ambient temperatures of these organisms (summarized in **Table 3.1**).

Table 3.1: Comparing the proportion of tRNAs with low ΔG values from six species.

Based on SI Table 3.1, the calculated average ΔG of the 250 tRNAs investigated is -33.48 kcal/mol. The number and percentage of tRNAs from each species with ΔG values less than the average ΔG are reported.

Species	total tRNAs	# of tRNAs	% of tRNAs	%GC Content		
		with less than average ΔG (more stable)		Minimum	Average	Maximum
H. thermophilus	44	33	75.00%	59%	66%	72%
A. aeolicus	42	38	90.48%	64%	69%	74%
T. thermophilus	45	33	73.33%	59%	66%	74%
B. subtilis	34	7	20.59%	49%	59%	67%
S. cerevisiae	42	1	2.38%	47%	56%	67%
E. coli	43	11	25.58%	49%	59%	68%

The tRNAs with the lowest free energies often have either a 6 bp D stem, 4 nt D loop, and large variable loop or a 3 bp D stem, 11 nt D loop, and large variable loop. Rarely do the tRNAs have a large variable loop without one of these pairs of conditions for the D stem and loop. Although *E. coli* and *B. subtilis* both have tRNAs that fit these criteria, there is a greater number of tRNAs that fit these criteria in *H. thermophilus*, *A. aeolicus*, and *T. thermophilus*.

A chart comparing the percentage of tRNAs from each organism with secondary structures larger than, equal to, or greater than the mode was also compiled. This helps to visualize that secondary structure elements with variable lengths, such as the acceptor stem, D stem, D loop, and variable loop, are varied in all six species examined. *A. aeolicus* and *H. thermophilus* do not have any obvious outliers compared to the other organisms.

3.8 Thermophilic tRNA thermal stability

Although tRNAs from thermophiles predictably have higher calculated ΔG values, there is little experimental data available about the behavior of thermophilic tRNAs compared to the well-characterized tRNAs from humans, *A. thaliana*, *E. coli*, and other species. To this end, UV melting experiments may be used to determine the melting temperature of the tertiary structure of precursor tRNAs used in the kinetic assays previously described. In general, tRNAs exhibit multiple unfolding events during denaturation that can be monitored via changes in absorbance at 260 nm, with the first major transition corresponding to tertiary structure (17, 18). Because previously characterized tRNAs, such as mitochondrial tRNAs, are known to unfold at temperatures far below the ambient temperatures of thermophilic organisms, it is possible that the thermophilic tRNAs may not adopt the same conformation at higher temperatures, limiting the physiological relevance of the kinetic assays performed in this work. UV melting experiments may be useful to determine if this is the case.

I have started preliminary experiments with UV melting using Ht ptRK5, Tt ptRG14, and Sc ptRF9. Buffer selection is important to ensure that the pH of the sample does not vary with temperature. Sodium cacodylate was selected as it has a low $\Delta pK_a/\Delta T$ value and is suitable for the pH range used. Initial experiments using a stir bar in the cuvette to ensure consistent heating throughout the sample had too much noise to be interpreted, perhaps due to the vortex created by

the stir bar or by providing a nucleation site for boiling. Exclusion of the stir bar led to the graphs

shown in **Figure 3.6**, **Figure 3.7**, and **Figure 3.8**. Although the data during the heating phase of

the experiment has too much noise to interpret reliably, the results can be seen in panels A-C of

Figure 3.6, **Figure 3.7**, and **Figure 3.8**. The cooling phase was able to be analyzed with more

certainty. From this preliminary data, Ht ptRK5 has melting transitions at 44 and 64 °C, Tt ptRG14

has melting transitions at 43 and 69 °C, and Sc ptRF9 has one melting transition at 52 °C. The

melting transition obtained for Sc ptRF9 is consistent with previously published studies which

identified melting transitions at 57 °C or ~55 °C under similar buffer conditions, providing

confidence in the conditions used here (26, 27). The T_M determined during the heating and cooling

phases vary by ~10 °C, which is consistent with previous studies that indicate RNA refolding

occurs slower than denaturation (28). Moving forward, it will be necessary to optimize the

experiment to obtain melting data during the heating phase, perhaps by using a more gradual

temperature ramp, and repeat the experiments in triplicate.
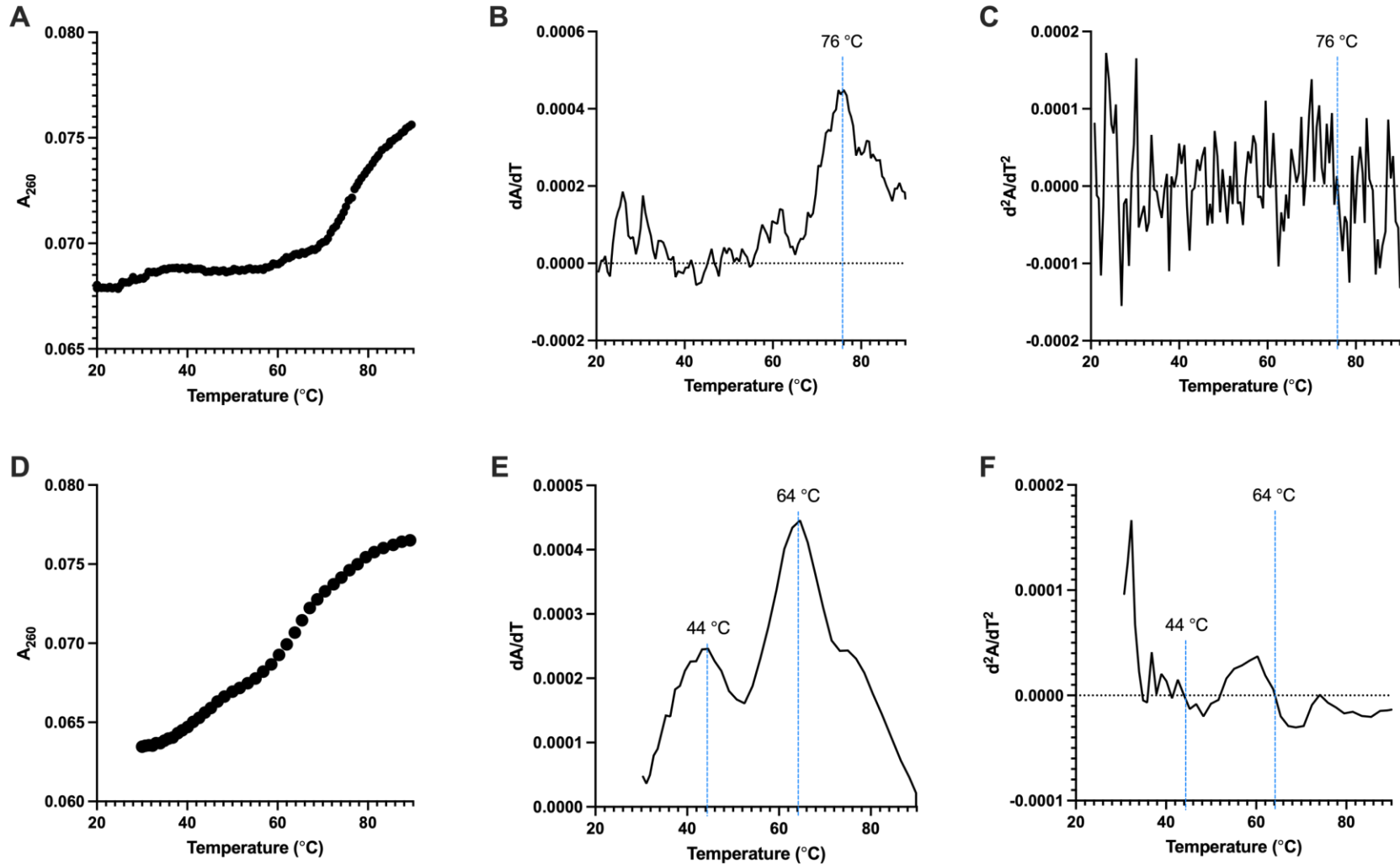
Figure 3.6: Thermal denaturation of *H. thermophilus* pre-tRNA[Lys 5:0].

A. and D. Absorbance readings of Ht ptRK5 as the temperature **A.** increased from 20-90 °C and **D.** decreased from 90 °C to 30 °C. **B. and E.** are first derivative plots of the data in **A.** and **D.**, respectively. **C. and F.** are second derivative plots of the data in **A.** and **D.**, respectively. The T_M of the sample is indicated with a dotted blue line based on local maxima in the first derivative plots and x-intercepts in the second derivative plots.
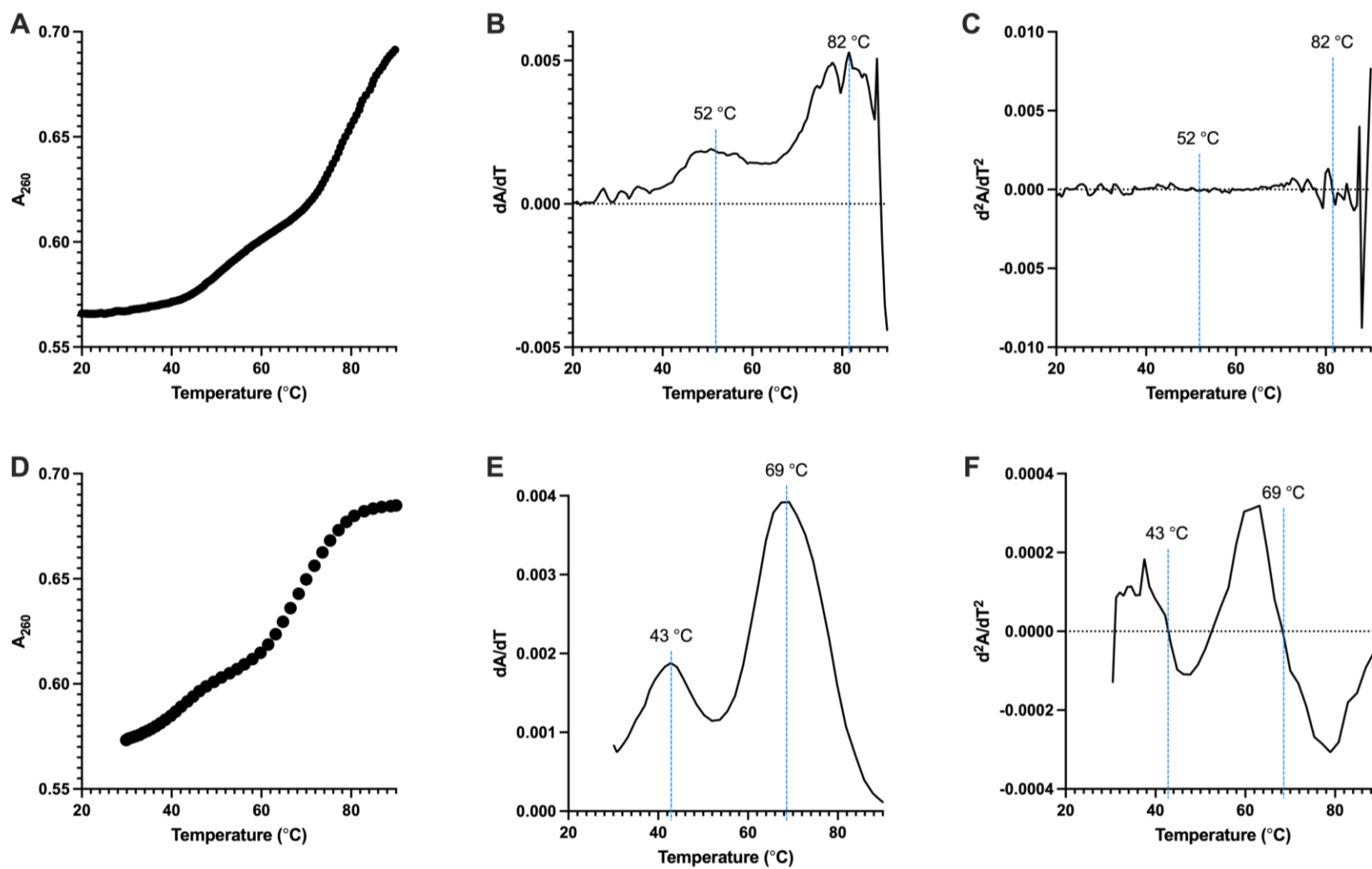
Figure 3.7: Thermal denaturation of *T. thermophilus* pre-tRNA^{Gly 14:0}.

A. and D. Absorbance readings of Tt ptRG14 as the temperature **A.** increased from 20-90 °C and **D.** decreased from 90 °C to 30 °C. **B. and E.** are first derivative plots of the data in **A.** and **D.**, respectively. **C. and F.** are second derivative plots of the data in **A.** and **D.**, respectively. The T_M of the sample is indicated with a dotted blue line based on local maxima in the first derivative plots and x-intercepts in the second derivative plots.
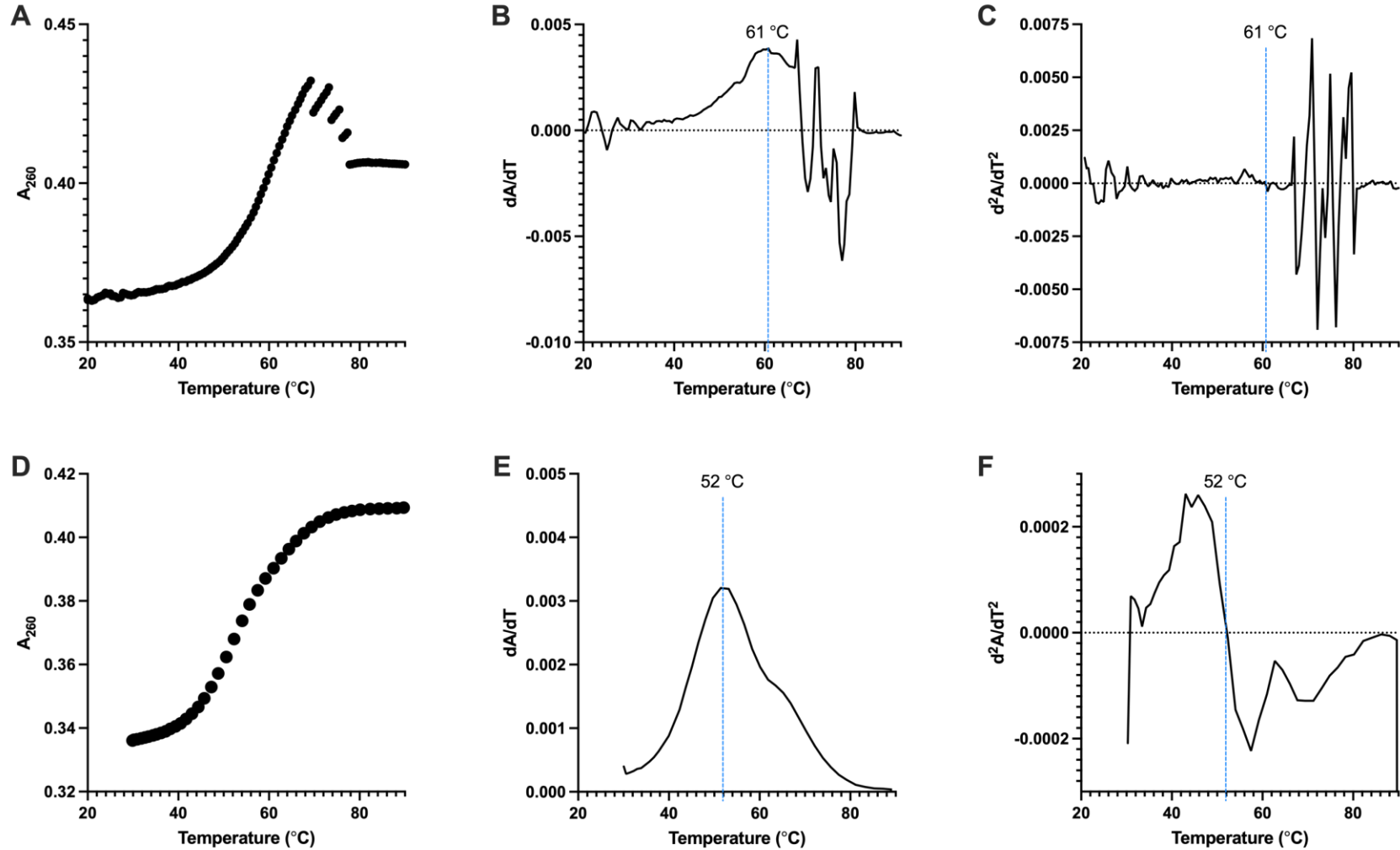
Figure 3.8: Thermal denaturation of S. cerevisiae pre-tRNA$^{Phe\ 9:0}$.

A. and D. Absorbance readings of Sc ptRF9 as the temperature **A.** increased from 20-90 °C and **D.** decreased from 90 °C to 30 °C. **B. and E.** are first derivative plots of the data in **A.** and **D.**, respectively. **C. and F.** are second derivative plots of the data in **A.** and **D.**, respectively. The T_M of the sample is indicated with a dotted blue line based on local maxima in the first derivative plots and x-intercepts in the second derivative plots.

3.9 Conclusions

In addition to our structural characterization of HARP, we investigated the single-turnover activity of Hth1307 using native and non-native substrates. Our experiments indicate a strong preference for the native Ht ptRK substrate, with activity towards native Ht ptRK[5] with the 3' CCA tail being 10-fold lower and activity towards non-native tRNAs being 60-fold lower. The disparity in observed rate constants between the different substrates used here is unlikely to be due to the phosphodiester cleavage and is instead likely due to slow binding of non-native substrates or a protein conformational change associated with substrate binding. Our attempts to characterize the binding affinity of Hth1307 for different substrates using electrophoretic mobility shift assays (data not shown) were complicated by the appearance of multiple bands, indicative of either multiple tRNAs binding to one oligomer or multiple oligomeric states of Hth1307 binding to tRNA. Regardless, the substrate specificity displayed by Hth1307 is uncommon in PRORP enzymes, which are generally considered more promiscuous and have similar activities for native and non-native substrates (7, 16, 18).

To investigate potential causes of the specificity of Hth1307, the secondary structure features and calculated ΔG values of tRNAs from six organisms were compared. Overall, thermophilic tRNAs have higher GC-content and lower ΔG values than tRNAs from other model organisms. However, they do not seem to have any remarkably different structural elements that may contribute to the specificity observed. Because HARPs are found in thermophiles, UV melting profiles were obtained for two thermophilic and one yeast pre-tRNA to determine if thermophilic tRNAs may adopt different conformations at higher temperatures. *T. thermophilus* was discovered in a hot spring at 80 °C and grows optimally at 65-72 °C (29), while *H. thermophilus* grows optimally at 70-75 °C (30). Both Tt ptRG14 and Ht ptRK5 have The first melting transitions of

both Tt ptRG14 and Ht ptRK5 are below the optimal growing temperatures of their respective organisms and higher than the temperatures used in assays so far (**SI Table 3.1** and **SI Table 3.2**), suggesting the tertiary structure of the tRNAs may be altered under physiological conditions. As such, it will be necessary to investigate the activity and structure of thermophilic tRNAs and HARPs under broader conditions in the future.

A major gap in our understanding of protein-only RNase P enzymes is the lack of data to make direct comparisons across species and classes. As shown in **SI Table 3.1** and **SI Table 3.2**, which compile an extensive list of kinetic and thermodynamic experiments with PRORPs, variations in the conditions used to obtain kinetic and thermodynamic values (such as k_{obs}, K_D, K_M, and k_{cat}) include differences in salt and magnesium concentrations, pH values, temperature, and protein and substrate concentrations and ratios. All the available studies use a wide array of substrates, many of which are non-native substrates and have varied 5' leader and 3' trailer lengths, which affect K_D and k_{obs} values. There is also little investigation into the optimal conditions for HARP enzymes, potentially due to a lack of information about the intracellular conditions across organisms leading to a limited idea of the range of conditions that should be examined.

In the work presented here, Hth1307 single-turnover activity (k_{obs}) for native substrates ranged 0.30-0.55 min^{-1}, which is similar to that of Aq880 with the non-native substrate Tt ptRG^{14} (0.62-2.06 min^{-1}), MRPP3 with mitochondrial pre-tRNAs (0.03-0.80 min^{-1}), At PRORP1 with non-native substrates (0.05-0.69 min^{-1}), and At PRORP2 with non-native substrates (0.03-0.31 min^{-1}) (7, 10, 11, 15, 18). Despite the variability of experimental conditions, the catalytic mechanism is likely conserved across all types of PRORP enzymes as evidenced by their conserved active sites (**Figure 2.9**) and similar kinetic activities.

SI Table 3.1: Comparison of PRORP kinetics data from available articles.

Protein	[Protein]	Pre-tRNA Substrate*	[RNA]	Native substrate?	[Mg]	[NaCl]	Temp (°C)	k_{obs} or k_{react} (min^{-1})	Reference
*Hth*1307	1 µM	*Ht* ptRK$^{5:0}$	2 nM	Yes	1 mM	150 mM	37	0.58	(Wilhelm et al., 2023)
		Ht ptRK$^{10:0}$						0.31	
		Ht ptRK$^{14:0}$						0.49	
		Ht ptRK$^{20:0}$						0.36	
		Ht ptRK^{5CCA}						0.05	
		Tt ptRG$^{14:0}$		No				0.008	
		Sc ptRF$^{9:0}$						0.006	
		Bs ptRD^{5CCA}						0.008	
*Aq*880	50 nM	*Tt* ptRG^{14}	5 nM	No	4.5 mM	20 mM	37	0.62	(Feyh et al., 2021)
	500 nM							2.06	
*Aq*880 Δ181-191	50 nM							0.22	
	500 nM							1.41	
*Aq*880 Δ181-191	500 nM							4×10^{-3}	
*Aq*880 Δ179-191	500 nM							7×10^{-3}	
*Aq*880 Δ177-191	500 nM							n.d.	
*Aq*880 R125A	50 nM							15×10^{-4}	
	500 nM							16.1×10^{-2}	
*Aq*880 R129A	500 nM							n.d.	
*Aq*880 R125A/R129A	500 nM							n.d.	
*Aq*880 K119A/R123A/R125A/K127A/R129A	500 nM							n.d.	
Aq880	varied	*Tt* ptRG^{14}	1 nM	No	4.5 mM	20 mM	37	1.43	(Nickel et al., 2017)
At PRORP1 ΔN76	2 µM	*At* ptRC$^{5:0}$	<40 nM	Yes	10 mM	100 mM NH$_4$OAc	37	1.2 ± 0.3	(Chen et al., 2016)
At PRORP1 ΔN76 K101A								1.2 ± 0.4	

Protein	[Protein]	Pre-tRNA Substrate*	[RNA]	Native substrate?	[Mg]	[NaCl]	Temp (°C)	k_{obs} or k_{react} (min^{-1})	Reference
At PRORP1 ΔN76 K109A								1.3 ± 0.3	
At PRORP1 ΔN76 K439A								0.4 ± 0.1	
At PRORP1 ΔN76 K460A								1.3 ± 0.1	
At PRORP1 ΔN76 K101A/K109A								2.5 ± 0.1	
At PRORP1 ΔN76 K101A/K439A								0.5 ± 0.1	
At PRORP1 ΔN76 K109A/K439A								0.5 ± 0.1	
At PRORP1	5 µM	*Bs* ptRD$^{1:0}$	30 nM	No	1 mM	150 mM	25	0.078	(Howard et al., 2016)
		Bs ptRD$^{2:0}$						0.15	
		Bs ptRD$^{3:0}$						0.032	
		Bs ptRD$^{4:0}$		No				0.025	
		Bs ptRD$^{5:0}$						0.025	
		Bs ptRD$^{10:0}$						0.025	
		Bs ptRD$^{14:0}$						0.02	
		At mito-ptRC$^{5:0}$		Yes				0.037	
		At chlor-ptRF$^{5:0}$						0.035	
		At nuc-ptRC$^{5:0}$		No				0.037	
		At nuc-ptRF$^{5:0}$						0.078	
At PRORP2		*At* mito-ptRC$^{5:0}$		No				0.013	
		At chlor-ptRF$^{5:0}$						0.018	
		At nuc-ptRC$^{5:0}$		Yes				0.027	
		At nuc-ptRF$^{5:0}$						0.035	
At PRORP3		*At* mito-ptRC$^{5:0}$		No				0.023	
		At chlor-ptRF$^{5:0}$						0.023	
		At nuc-ptRC$^{5:0}$		Yes				0.03	
		At nuc-ptRF$^{5:0}$						0.072	

Protein	[Protein]	Pre-tRNA Substrate*	[RNA]	Native substrate?	[Mg]	[NaCl]	Temp (°C)	k_{obs} or k_{react} (min⁻¹)	Reference
At PRORP1 ΔN76	5 µM	*At* mito-ptRC$^{5:0}$	20 nM	Yes	1 mM	150 mM	25	0.04	(Howard et al., 2015)
		Bs ptRD$^{5:0}$		No				0.03	
At PRORP1 ΔN76	500 nM	*At* mito-ptRC$^{53:24}$	25 pM	Yes	1 mM	150 mM	25	1.4	(Howard et al., 2012)
At PRORP2	5 µM	*At* nuc-ptRG$^{8:1}$	50 nM	Yes	1 mM	150 mM	25	1.1	(Karasik et al., 2016)
		At nuc-ptRG$^{13:1}$						0.7	
		At nuc-ptRG$^{23:1}$						0.7	
		At nuc-ptRG$^{23:5}$						1	
		At nuc-ptRG$^{23:10}$						0.7	
Hs PRORP	0.1-2 µM	*Hs* mito-ptRL$^{(UUR)6:0}$	20 nM	Yes	1 mM	150 mM	28	0.41	(Karasik et al., 2019)
		Hs mito-ptRI$^{7:0}$						0.8	
		Hs mito-ptRM$^{6:0}$						0.45	
	1.5 µM	*Hs* mito-ptRV$^{7:0}$						0.3	
		Hs mito-ptRS$^{(UCN)7:0}$						0.06	
		Hs mito-ptRS$^{(AGY)7:0}$						0.06	
At PRORP1		*Hs* mito-ptRL$^{(UUR)6:0}$		No				0.06	
		Hs mito-ptRI$^{7:0}$						0.05	
		Hs mito-ptRM$^{6:0}$						0.09	
		Hs mito-ptRV$^{7:0}$						0.08	
		Hs mito-ptRS$^{(UCN)7:0}$						0.69	
		Hs mito-ptRS$^{(AGY)7:0}$						0.09	
At PRORP2		*Hs* mito-ptRL$^{(UUR)6:0}$						0.04	
		Hs mito-ptRI$^{7:0}$						0.03	
		Hs mito-ptRM$^{6:0}$						0.03	
		Hs mito-ptRV$^{7:0}$						0.03	
		Hs mito-ptRS$^{(UCN)7:0}$						0.31	
		Hs mito-ptRS$^{(AGY)7:0}$						0.04	
At PRORP1	2 µM	*S. sp.* PCC6803 ptRQ	40 nM	No	1 mM	150 mM	4	0.023	(Chen et al., 2019)
							15	0.17	
							25	0.8	
							37	2.3	
							42	3.1	
							45	2.6	

Protein	[Protein]	Pre-tRNA Substrate*	[RNA]	Native substrate?	[Mg]	[NaCl]	Temp (°C)	k_{obs} or k_{react} (min^{-1})	Reference
At PRORP2							4	0.053	
							15	0.37	
							25	1	
							37	3.3	
							42	1.6	
							45	0	
At PRORP3							4	0.063	
							15	0.42	
							25	1.2	
							37	3.7	
							42	4.8	
							45	0	
At PRORP1	2 µM	*S. sp.* PCC6803 ptRQ$^{10:CCA}$ (D+1)	40 nM	No	1 mM	150 mM	25	0.8	
							37	2.4	
							42	3.2	
		S. sp. PCC6803 ptRQ$^{10:CCA}$ (D+6)					25	0.4	
							37	1.6	
							42	1.5	
		S. sp. PCC6803 ptRQ$^{10:CCA}$ ΔAC					25	0.7	
							37	2.3	
							42	3.1	
		At chloro-ptRR$^{5:5}$		Yes			25	0.06	
							37	0.2	
							42	0.14	
At PRORP2		S. sp. PCC6803 ptRQ$^{10:CCA}$ (D+1)		No			25	0.8	
							37	2.9	
							42	1.2	
		S. sp. PCC6803 ptRQ$^{10:CCA}$ (D+6)					25	0.2	
							37	0.3	
							42	0	
		S. sp. PCC6803 ptRQ$^{10:CCA}$ ΔAC					25	0.7	
							37	2.4	
							42	ND	

Protein	[Protein]	Pre-tRNA Substrate*	[RNA]	Native substrate?	[Mg]	[NaCl]	Temp (°C)	k_{obs} or k_{react} (min^{-1})	Reference
At PRORP3		*At* chloro-ptRR$^{5:5}$					25	0.06	
							37	0.21	
							42	0	
		S. sp. PCC6803 ptRQ$^{10:CCA}$ (D+1)					25	1	
							37	3.5	
							42	2.1	
		S. sp. PCC6803 ptRQ$^{10:CCA}$ (D+6)					25	0.4	
							37	0.7	
							42	0	
		S. sp. PCC6803 ptRQ$^{10:CCA}$ ΔAC					25	0.9	
							37	3.5	
							42	1.2	
		At chloro-ptRR$^{5:5}$					25	0.07	
							37	0.44	
							42	0	
At PRORP3	225 nM	*Tt* ptRG^{7}	trace amounts	No	10 mM	20 mM	20	1.52	(Walczyk et al., 2016)
	varied				4.5 mM			1.11	
At PRORP1 ΔN76	2.5-22.5 µM	*Bs* ptRD^{5}	30 nM	No	20 mM	330 mM	25	2.62	(Teramoto et al., 2020)
Hs (mt)RNase P	50 nM PRORP; 200 nM TRMT10C-SDR5C1	*Hs* ptRA	0.2 nM	Yes	3 mM	20 mM	30	1.54	(Vilardo et al., 2023)
		Hs ptRM			4.5 mM			1.18	
		Hs ptRI			3 mM			8.27	
		Hs ptRK			4.5 mM			0.82	
		Hs ptRV			4.5 mM			0.77	

* Pre-tRNA names are abbreviated by the following naming convention: Bs ptRD5:1 is *Bacillus subtilis* pre-tRNA^{Asp} with a 5-nt leader sequence and a 1-nt trailer sequence. A 0 for the trailer sequence indicates that there is only a discriminator base. Abbreviations used: *Aq*, *Aquifex aeolicus; Ht, Hydrogenobacter thermophilus; Tt, Thermus thermophilus*; Sc, Saccharomyces cerevisiae; *Bs*, *Bacillus subtilis*; *At*, *Arabidopsis thaliana; S. sp.* PCC6803, *Synechocystis sp*. PCC6803; *Hs*, *Homo sapiens;* mito, mitochondria; chloro, chloroplast; nuc, nuclear.

SI Table 3.2: Comparison of PRORP binding data from available articles.

Protein	[PRORP]	pre-tRNA Substrate*	[RNA]	Native Substrate?	[Ca]	[NaCl]	Temp (°C)	K_D (nM)	Reference
At PRORP1 ΔN76	0-20 µM	*At* mito-ptRC	20 nM	Yes	1 mM	100 mM	Room Temp	700± 0.1	(Howard et al., 2012)
At PRORP1 ΔN76								700 ± 80	
At PRORP1 ΔN76 H498A								680 ± 150	
At PRORP1 ΔN76 H498Q	0.005-5 µM	*At* mito-ptRC$^{5:1}$	20 nM	Yes	1 mM	300 mM	25	710 ± 60	(Howard et al., 2015)
At PRORP1 ΔN76 H438A								700 ± 120	
		Bs ptRD$^{0:0}$						3400 ± 400	
		Bs ptRD$^{1:0}$						150 ± 60	
		Bs ptRD$^{2:0}$						310 ±20	
		Bs ptRD$^{3:0}$						140 ± 40	
At PRORP1 ΔN76	0.005-20 µM	*Bs* ptRD$^{4:0}$		No	1 mM	300 mM	Room Temp	150 ± 40	(Howard et al., 2016)
		Bs ptRD$^{5:0}$	20 nM					190 ± 60	
		Bs ptRD$^{10:0}$						100 ± 50	
		Bs ptRD$^{14:0}$						100 ± 30	
		At mito-ptRC$^{5:0}$		Yes				510 ± 120	
At PRORP1 ΔN76	0.005-20 µM	*At* chloro-ptRF$^{5:0}$			1 mM	300 mM	Room Temp	60 ± 10	
		At nuc-ptRC$^{5:0}$		No				2300 ± 300	
		At nuc-ptRF$^{5:0}$						330 ± 60	

Protein	[PRORP]	pre-tRNA Substrate*	[RNA]	Native Substrate?	[Ca]	[NaCl]	Temp (°C)	K_D (nM)	Reference
At PRORP2		*At* mito-ptRC$^{5:0}$		No				350 ± 70	
		At chloro-ptRF$^{5:0}$						140 ± 10	
		At nuc-ptRC$^{5:0}$		Yes				6100 ± 2100	
		At nuc-ptRF$^{5:0}$						350 ± 40	
At PRORP3 ΔN9		*At* mito-ptRC$^{5:0}$		No				300 ± 70	
		At chloro-ptRF$^{5:0}$						220 ± 30	
		At nuc-ptRC$^{5:0}$		Yes				1500 ± 200	
		At nuc-ptRF$^{5:0}$						380 ± 50	
At PRORP2	0-1000 nM	*At* nuc-ptRG$^{23:10}$	20 nM	yes	6 mM	150 mM	25	118 ± 26	(Karasik et al., 2016)
		At nuc-ptRG$^{23:5}$						52 ± 12	
		At nuc-ptRG$^{23:1}$						17 ± 5	
		At nuc-ptRG$^{13:1}$						6 ± 1	
		At nuc-ptRG$^{8:1}$						3 ± 1	
At PRORP1 ΔN76		*At* ptRC$^{5:23}$	20 nM	Yes	10 mM	100 mM NH$_4$OAc	25	250 ± 34	(Chen et al., 2016)
At PRORP1 ΔN76 K101A								859 ±159	
At PRORP1 ΔN76 K109A								389 ± 35	
At PRORP1 ΔN76 K439A								609 ± 70	
At PRORP1 ΔN76 K460A								377± 23	

Protein	[PRORP]	pre-tRNA Substrate*	[RNA]	Native Substrate?	[Ca]	[NaCl]	Temp (°C)	K_D (nM)	Reference
At PRORP1 ΔN76 K101A/K109A								1624 ±344	
At PRORP1 ΔN76 K101A/K439A								888 ±127	
At PRORP1 ΔN76 K109A/K439A								763 ± 92	
At PRORP1 ΔN76		*At* ptRC[5:5]						77 ± 19	
At PRORP1 ΔN76	≥ 3x KD	*At* mito-pTRC[5:0]	≤ 5x KD	Yes	20 mM	330 mM	28	1330 ± 120	(Klemm et al., 2017)
		Bs pTRD[5:0]		No				155 ± 20	
		Bs pTRD[1:0]						600 ± 50	
		Bs pTRD[0:0]						25400 ± 6100	
At PRORP2		*At* nuc-pTRG[6:0]		Yes	6 mM	330 mM		470 ± 50	
		Bs pTRD[5:0]		No				80 ± 9	
		Bs pTRD[0:0]						17200 ± 1100	
Hs (mt)RNase P	150 nM TRMT10C-SDR5C1 ΔN39; 0-50 µM Hs PRORP ΔN95	*Hs* mito-ptRL[(UUR) 6:0]	20 nM	yes	6 mM	150 mM	28	50.6 ± 5.3	(Karasik et al., 2019)
		Hs mito-ptRI[7:0]						13.5 ± 2.3	
		Hs mito-ptRV[7:0]						4124 ± 560	
		Hs mito-ptRM[6:0]						707 ±168	
		Hs mito-ptRS[(UCN)7:0]						3556 ± 540	
		Hs mito-ptRS[(AGY)7:0]						12,660 ± 1312	
At PRORP1	Not Reported	*Hs* mito-ptRL[(UUR) 6:0]	20 nM	No	6 mM	150 mM	28	883 ± 174	
		Hs mito-ptRI[7:0]						4105 ± 1400	

Protein	[PRORP]	pre-tRNA Substrate*	[RNA]	Native Substrate?	[Ca]	[NaCl]	Temp (°C)	K_D (nM)	Reference
		Hs mito-ptRV$^{7:0}$						>> 10000	
		Hs mito-ptRM$^{6:0}$						> 5000	
		Hs mito-ptRS$^{(UCN)7:0}$						123 ±23	
		Hs mito-ptRS$^{(AGY)7:0}$						>> 10000	
At PRORP2	Not Reported	Hs mito-ptRL$^{(UUR)\ 6:0}$	20 nM	No	6 mM	150 mM	28	349 ± 40	
		Hs mito-ptRI$^{7:0}$						1155 ± 167	
		Hs mito-ptRV$^{7:0}$						4300 ± 690	
		Hs mito-ptRM$^{6:0}$						1596 ± 163	
		Hs mito-ptRS$^{(UCN)7:0}$						59 ± 10	
		Hs mito-ptRS$^{(AGY)7:0}$						> 5000	
Hs (mt)RNase P	150 nM TRMT10C-SDR5C1 ΔN39; 0-50 µM *Hs* PRORP ΔN95	Hs ptRV$^{7:0}$	20 nM	Yes	6 mM	150 mM	28	1626 ± 375	
		Hs ptRV$^{7:0}$ w/ 25 µM SAM						101 ± 13	
		Hs ptRV$^{7:0/C9}$	20 nM	No				3313 ± 800	
		Hs ptRV$^{7:0/C9}$ w/ 25 µM SAM						80 ± 23	
At PRORP1 ΔN76	Not Reported	*S. sp.* PCC6803 ptRQ$^{10:CCA}$	20 nM	Yes	1 mM	150 mM	22-25	152 ± 62	(Chen et al., 2019)
		S. sp. PCC6803 ptRQ$^{10:CCA}$ (D+1)						86 ± 16	
		S. sp. PCC6803 ptRQ$^{10:CCA}$ (D+6)						676 ± 217	

Protein	[PRORP]	pre-tRNA Substrate*	[RNA]	Native Substrate?	[Ca]	[NaCl]	Temp (°C)	K_D (nM)	Reference
		S. sp. PCC6803 ptRQ$^{10:CCA}$ ΔAC						175 ± 46	
		At ptRR$^{5:5}$						887 ± 169	
At PRORP2	Not Reported	*S. sp.* PCC6803 ptRQ$^{10:CCA}$						276 ± 82	
		S. sp. PCC6803 ptRQ$^{10:CCA}$ (D+1)						2113 ± 982	
		S. sp. PCC6803 ptRQ$^{10:CCA}$ (D+6)						3224 ± 804	
		S. sp. PCC6803 ptRQ$^{10:CCA}$ ΔAC						2324 ± 1853	
		At ptRR$^{5:5}$						994 ± 207	
At PRORP3 ΔN9	Not Reported	*S. sp.* PCC6803 ptRQ$^{10:CCA}$						1849 ± 309	
		S. sp. PCC6803 ptRQ$^{10:CCA}$ (D+1)						>5000	
		S. sp. PCC6803 ptRQ$^{10:CCA}$ (D+6)						2507 ± 314	
		S. sp. PCC6803 ptRQ$^{10:CCA}$ ΔAC						>5000	
		At ptRR$^{5:5}$						1008 ± 202	
At PRORP1 ΔN76	0.009-20 μM	*Sc* ptRF$^{0:CCA}$	20 nM	No	20 mM	330 mM	25 ± 1	627 ± 55	(Teramoto et al., 2020)
At PRORP1 ΔN83								281 ± 44	

Protein	[PRORP]	pre-tRNA Substrate*	[RNA]	Native Substrate?	[Ca]	[NaCl]	Temp (°C)	K_D (nM)	Reference
At PRORP1 ΔN83 ΔPPR2 loop								317 ± 92	
At PRORP1 ΔN86								3405 ± 338	
At PRORP1 ΔN83		*At* mito-ptRC		Yes				257 ± 54	
At PRORP1 ΔN76		*Bs* ptRD5		No				390 ± 30	
Hs(mt)PRORP	25-1000 nM	*Hs* mito-ptRA	Not Reported	Yes	Not Reported	20 mM	21	14.8	(Vilardo et al., 2023)
		Hs mito-ptRH						23.0	
		Hs mito-ptRK						26.3	
		Hs mito-ptRV						17.3	
Pb HARP		*E.coli* ptRH	1 µM	No	0 mM	200 mM	On ice	4.3	(Y. Li et al., 2022)
Pb HARP R116A								4.2	
Pb HARP R123A	1-10 µM							3.8	
Pb HARP T135A								4.1	
Pb HARP R138A								3.8	
Pb HARP R142A								4.5	
Pb HARP ΔN								8.8	
Pb HARP ΔN/R116A	1-20 µM							9.1	
Pb HARP ΔN/R123A								9.7	
Pb HARP ΔN/R138A								15.6	

Protein	[PRORP]	pre-tRNA Substrate*	[RNA]	Native Substrate?	[Ca]	[NaCl]	Temp (°C)	K_D (nM)	Reference
Pb HARP ΔN/R142A								9.7	
Pb HARP ΔN/R138A/R142A								>20.0	

* Pre-tRNA names are abbreviated by the following naming convention: *Bs* ptRD5:1 is *Bacillus subtilis* pre-tRNA[Asp] with a 5-nt leader sequence and a 1-nt trailer sequence. A 0 for the trailer sequence indicates that there is only a discriminator base. Abbreviations used: *S. sp.* PCC6803, *Synechocystis sp.* PCC 6803; *Bs*, *Bacillus subtilis*; *At*, *Arabidopsis thaliana*; *Pb*, *Planctomycetes bacterium*; *Hs*, *Homo sapiens*; mito, mitochondria; chloro, chloroplast; nuc, nuclear.

3.10 References

1. Holzmann, J., Frank, P., Löffler, E., Bennett, K. L., Gerner, C., and Rossmanith, W. (2008) RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell*. **135**, 462–474

2. Gobert, A., Gutmann, B., Taschner, A., Gößringer, M., Holzmann, J., Hartmann, R. K., Rossmanith, W., and Giegé, P. (2010) A single Arabidopsis organellar protein has RNase P activity. *Nat Struct Mol Biol*. **17**, 740–744

3. Taschner, A., Weber, C., Buzet, A., Hartmann, R. K., Hartig, A., and Rossmanith, W. (2012) Nuclear RNase P of Trypanosoma brucei: A Single Protein in Place of the Multicomponent RNA-Protein Complex. *Cell Reports*. **2**, 19–25

4. Bhatta, A., Dienemann, C., Cramer, P., and Hillen, H. S. (2021) Structural basis of RNA processing by human mitochondrial RNase P. *Nat Struct Mol Biol*. **28**, 713–723

5. Howard, M. J., Lim, W. H., Fierke, C. A., and Koutmos, M. (2012) Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5′ processing. *Proceedings of the National Academy of Sciences*. **109**, 16149–16154

6. Teramoto, T., Kaitany, K. J., Kakuta, Y., Kimura, M., Fierke, C. A., and Hall, T. M. T. (2020) Pentatricopeptide repeats of protein-only RNase P use a distinct mode to recognize conserved bases and structural elements of pre-tRNA. *Nucleic Acids Res*. **48**, 11815–11826

7. Howard, M. J., Karasik, A., Klemm, B. P., Mei, C., Shanmuganathan, A., Fierke, C. A., and Koutmos, M. (2016) Differential substrate recognition by isozymes of plant protein-only Ribonuclease P. *RNA*. **22**, 782–792

8. Gobert, A., Quan, Y., Arrivé, M., Waltz, F., Da Silva, N., Jomat, L., Cohen, M., Jupin, I., and Giegé, P. (2021) Towards plant resistance to viruses using protein-only RNase P. *Nat Commun*. **12**, 1007

9. Dreher, T. W. (2009) Role of tRNA-like structures in controlling plant virus replication. *Virus research*. **139**, 217

10. Nickel, A. I., Wäber, N. B., Gößringer, M., Lechner, M., Linne, U., Toth, U., Rossmanith, W., and Hartmann, R. K. (2017) Minimal and RNA-free RNase P in Aquifex aeolicus. *Proceedings of the National Academy of Sciences*. **114**, 11121–11126

11. Feyh, R., Waeber, N. B., Prinz, S., Giammarinaro, P. I., Bange, G., Hochberg, G., Hartmann, R. K., and Altegoer, F. (2021) Structure and mechanistic features of the prokaryotic minimal RNase P. *eLife*. **10**, e70160

12. Teramoto, T., Koyasu, T., Adachi, N., Kawasaki, M., Moriya, T., Numata, T., Senda, T., and Kakuta, Y. (2021) Minimal protein-only RNase P structure reveals insights into tRNA precursor recognition and catalysis. *J Biol Chem*. **297**, 101028

13. Li, Y., Su, S., Gao, Y., Lu, G., Liu, H., Chen, X., Shao, Z., Zhang, Y., Shao, Q., Zhao, X., Yang, J., Cao, C., Lin, J., Ma, J., and Gan, J. (2022) Crystal structures and insights into precursor tRNA 5'-end processing by prokaryotic minimal protein-only RNase P. *Nat Commun*. **13**, 2290

14. Klemm, B. P., Karasik, A., Kaitany, K. J., Shanmuganathan, A., Henley, M. J., Thelen, A. Z., Dewar, A. J. L., Jackson, N. D., Koutmos, M., and Fierke, C. A. (2017) Molecular recognition of pre-tRNA by Arabidopsis protein-only Ribonuclease P. *RNA*. **23**, 1860–1873

15. Karasik, A., Shanmuganathan, A., Howard, M. J., Fierke, C. A., and Koutmos, M. (2016) Nuclear Protein-only ribonuclease P 2 Structure and Biochemical Characterization Provide Insight Into the Conserved Properties of tRNA 5′-end Processing Enzymes. *J Mol Biol*. **428**, 26–40

16. Howard, M. J., Klemm, B. P., and Fierke, C. A. (2015) Mechanistic Studies Reveal Similar Catalytic Strategies for Phosphodiester Bond Hydrolysis by Protein-only and RNA-dependent Ribonuclease P. *J Biol Chem*. **290**, 13454–13464

17. Karasik, A., Wilhelm, C. A., Fierke, C. A., and Koutmos, M. (2021) Disease-associated mutations in mitochondrial precursor tRNAs affect binding, m1R9 methylation, and tRNA processing by mtRNase P. *RNA*. **27**, 420–432

18. Karasik, A., Fierke, C. A., and Koutmos, M. (2019) Interplay between substrate recognition, 5' end tRNA processing and methylation activity of human mitochondrial RNase P. *RNA*. **25**, 1646–1660

19. Rueda, D., Hsieh, J., Day-Storms, J. J., Fierke, C. A., and Walter, N. G. (2005) The 5' leader of precursor tRNAAsp bound to the Bacillus subtilis RNase P holoenzyme has an extended conformation. *Biochemistry*. **44**, 16130–16139

20. Wu Meyers, N., Karasik, A., Kaitany, K., Fierke, C. A., and Koutmos, M. (2022) Gambogic acid and juglone inhibit RNase P through distinct mechanisms. *J Biol Chem*. **298**, 102683

21. Lan, P., Tan, M., Zhang, Y., Niu, S., Chen, J., Shi, S., Qiu, S., Wang, X., Peng, X., Cai, G., Cheng, H., Wu, J., Li, G., and Lei, M. (2018) Structural insight into precursor tRNA processing by yeast ribonuclease P. *Science*. **362**, eaat6678

22. Nakamura, Y., Gojobori, T., and Ikemura, T. (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res*. **28**, 292

23. Reiner, R., Alfiya-Mor, N., Berrebi-Demma, M., Wesolowski, D., Altman, S., and Jarrous, N. (2011) RNA binding properties of conserved protein subunits of human RNase P. *Nucleic Acids Research*. **39**, 5704–5714

24. Walczyk, D., Gößringer, M., Rossmanith, W., Zatsepin, T. S., Oretskaya, T. S., and Hartmann, R. K. (2016) Analysis of the Cleavage Mechanism by Protein-Only RNase P Using Precursor tRNA Substrates with Modifications at the Cleavage Site. *Journal of Molecular Biology*. **428**, 4917–4928

25. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., and Hofacker, I. L. (2008) The Vienna RNA Websuite. *Nucleic Acids Res*. **36**, W70–W74

26. Einert, T. R., and Netz, R. R. (2011) Theory for RNA Folding, Stretching, and Melting Including Loops and Salt. *Biophys J*. **100**, 2745–2753

27. Draper, D. E., Bukhman, Y. V., and Gluick, T. C. (2000) Thermal Methods for the Analysis of RNA Folding Pathways. *Current Protocols in Nucleic Acid Chemistry*. **2**, 11.3.1-11.3.13

28. Howe, C. P., Greetham, G. M., Procacci, B., Parker, A. W., and Hunt, N. T. (2022) Measuring RNA UNCG Tetraloop Refolding Dynamics Using Temperature-Jump/Drop Infrared Spectroscopy. *J. Phys. Chem. Lett.* **13**, 9171–9176

29. Oshima, T., and Imahori, K. (1974) Description of Thermus thermophilus (Yoshida and Oshima) comb. nov., a Nonsporulating Thermophilic Bacterium from a Japanese Thermal Spa. *International Journal of Systematic and Evolutionary Microbiology*. **24**, 102–112

30. Kawasumi, T., Igarashi, Y., Kodama, T., and Minoda, Y. (1984) Hydrogenobacter thermophilus gen. nov. , sp. nov. , an Extremely Thermophilic, Aerobic, Hydrogen-Oxidizing Bacterium. *International Journal of Systematic and Evolutionary Microbiology*. **34**, 5–10

Chapter 4 Conclusions and Future Directions

4.1 Conclusions

Ribonuclease P enzymes are essential in nearly all life, linked to a variety of neurological and cardiovascular diseases, and hold promise as a biological tool to combat viruses with tRNA-like structures. As such, it is important to determine the molecular architecture, substrate specificity, and molecular mechanism of RNase P enzymes. Complete lists of currently available protein-only RNase P structures, kinetic data, and binding data are presented in **SI Table 4.1**, **SI Table 3.1**, and **SI Table 3.2**, respectively, highlight the steady progress on this work over the last decade. In 2017, a novel class of protein-only RNase P enzymes was discovered in *Aquifex aeolicus* termed HARPs. This work integrates structural analysis with biochemical characterization to establish a foundation for understanding HARPs.

In **Chapter 2**, a combination of cryo-EM, x-ray crystallography, native mass spectrometry, and analytical size exclusion chromatography enabled the characterization of a novel tetradecameric state of HARP from *Hydrogenobacter thermophilus*. Alignment of this structure to other PRORPs shows a conserved active site, suggesting a shared catalytic mechanism. The active site of the HARP enzyme is positioned ~42 Å away from a region of positively charged residues that may serve as a substrate recognition site that would interact with the negatively charged phosphate backbone of the substrate pre-tRNA. From this data, a novel pre-tRNA binding model was proposed, lending insight into the substrate recognition strategy of this novel PRORP.

In **Chapter 3**, kinetic characterization of *H. thermophilus* HARP showed oligomerization is required for cleavage activity, supporting the proposed binding models. The enzyme activity is correlated with magnesium concentration, as seen with other PRORPs, further supporting a shared catalytic mechanism. When comparing the observed rate constant of *H. thermophilus* HARP with a variety of pre-tRNA substrates, the enzyme has similar levels of activity for a substrate with a 5- to 20-nt leader sequence, higher activity toward a substrate lacking the 3'-CCA tail used to aminoacylate the tRNA, and higher activity toward the native Ht ptRK5 compared to non-native Tt ptRG14, Sc ptRF9, and Bs ptRD5. Whether HARP displays a difference in binding affinity for these substrates and the reason for these differences in activity remain undetermined.

These findings support the idea that, despite a strikingly different molecular architecture, HARPs likely use a similar substrate recognition strategy and catalytic mechanism to other previously characterized PRORPs. However, several questions about HARP enzymes remain unanswered. The following section highlights these remaining questions and suggests key areas for future investigations.

4.2 Future Directions

4.2.1 Gaining insight into the physiologically relevant structure of HARPs

Because currently available structural data is collected at temperatures far lower than physiologically relevant temperatures, it is pertinent to examine the protein structure at higher temperatures. Circular dichroism (CD) is one of few techniques that offer this capability. CD takes advantage of the chiral nature of peptides by monitoring the difference in absorption of left-handed (ε_L) and right-handed (ε_R) circularly polarized light (1, 2). Measuring the CD spectrum of a protein sample as a function of temperature can be used to determine the melting temperature (T_M) of protein unfolding and monitor changes in secondary structure content (1–4). This could provide

insights into whether HARPs adopt similar structures in the structural data presented here and at the high temperatures of thermophilic organisms.

Another avenue to explore the physiologically relevant structure of HARP is cryo-electron tomography, which can be used to obtain the structure of protein complexes within cells (5–7). Although this technique uses cryogenic flash freezing similar to cryo-EM, this could allow the protein structure to be characterized within *A. aeolicus*, avoiding the variables introduced during protein purification procedures that may lead to reduced activity and different oligomeric states. It may even be possible to capture a structure of HARP complexed with tRNA using this technique.

Alternatively, atomic force microscopy (AFM) may be useful in monitoring conformational changes and substrate binding under near-physiological conditions (8, 9). Though not a technique currently used by our lab, AFM has previously been used to characterize protein-DNA complexes in DNA repair pathways (8) and could give valuable information about HARPs. Regardless of the technique used, one of the obvious future steps in characterizing HARPs is bridging the significant gap in our understanding of how the structure of HARP may change with temperature.

4.2.2 Characterizing the binding of pre-tRNA to HARPs

Although multiple HARP binding models have been proposed, they have yet to be confirmed. Solving the structure of a HARP•tRNA complex remains a challenge. Multiple groups have reported an an inability to isolate the HARP•tRNA complex using size exclusion chromatography (10). Crosslinking may be a valuable tool for obtaining a stable HARP•tRNA complex that can be used in structural studies. This could also provide information about the binding stoichiometry of the complex, which has also been stymied by difficulties isolating the complex. If the tRNA is susceptible to degradation during structure experiments, a substrate with

a synthetic backbone resistant to cleavage by RNases may be helpful in obtaining a stable complex (11, 12). If obtaining a structure of the HARP•tRNA complex remains challenging, or if more structural data is required, footprinting can provide information about protein-RNA interactions with single nucleotide resolution (13).

Binding affinity is another metric that would be useful in learning more about the interactions between HARP and substrate pre-tRNAs. As indicated in **Chapter 3**, it has been difficult to obtain reliable and decipherable data using traditional binding assays such as electrophoretic mobility shift assays and fluorescence polarization. Using MTO assays could at least provide some insight into the binding affinity of HARPs using K_M, although it is not directly equivalent to K_D and is discussed in further detail in the next section.

4.2.3 Using kinetics to explore optimal conditions for HARPs

To gain a more complete understanding of the activity of HARPs, kinetic assays can be performed under a wider variety of conditions than those tested here. For initial investigation, multiple turnover (MTO) assays can be used to determine the turnover number (k_{cat}) and Michaelis constant (K_M) of the reaction under the same buffer conditions used in the single turnover (STO) assays presented in **Chapter 3**. This initial foray into MTO assays would also provide a baseline for the following experiments.

Although multiple groups have reported structures of HARP coupled with kinetic data, there is no evidence that directly supports the assumption of two magnesium ions being used the catalytic mechanism. This assumption is based on the structure of the HARP active site residues being structurally homologous to those of human (mt)RNase P and *At* PRORPs, although none of the published HARP structures have two metal ions in the active site. In this work, STO assays were done at varied magnesium chloride concentrations, but increasing the concentration beyond

1 mM created difficulty capturing a complete time course showing conversion of substrate to product without the use of quench-flow kinetics. MTO assays performed with a variety of magnesium chloride concentrations would be a simple way to circumvent this difficulty and ascertain whether one or two magnesium ions are used in the active site.

Likewise, no work has been done to test the activity of HARPs at varied pH levels, salt concentrations, or temperatures, despite HARPs being found exclusively in thermophiles, which are known to thrive in a wide range of temperatures, salinity, and pH. This may be another opportune place to employ MTO kinetics instead of STO kinetics, as increasing the temperature and determining optimum salinity and pH will likely lead to further increased activity that would be difficult to monitor with STO assays. Varying these conditions would be particularly interesting as some organisms that have HARP have optimal growth at temperatures up to 85 °C, salt concentrations up to 1 M, and pH up to 9.0 (16–20), which are far beyond the conditions assayed so far and may alter the structure and dynamics of both the protein and the pre-tRNA substrates.

4.3 Closing thoughts

Since the discovery of protein-only RNase P, the evolutionary pathways and selective pressures leading to the coexistence of a ribozyme-based and protein-only RNase P have been of great interest. Although some form of the ribozyme RNase P has been retained throughout most living organisms, some have evolved to use only PRORPs. Interestingly, it has been shown that both RNA- and protein-based RNase P use the same catalytic strategy, coordinating two metal ions (Mg^{2+}) to perform phosphodiester bond hydrolysis. Key aspects of substrate recognition also seem to be conserved, as all forms of RNase P described use a molecular ruler method to bind the elbow region of the pre-tRNA and to position the cleavage site of the substrate in the active site of the enzyme. As the only example of a shared biological role in RNA- and protein-based catalysis,

RNase P may be a remnant of an evolutionary shift from RNA-based to protein-based catalysis and life, as predicted by the RNA World theory.

One of the key remaining questions in the field is, what are the evolutionary advantages of this split between RNA-based and protein-only RNase P? The RNP complex exhibits significantly higher cleavage activity than PRORP (14), and though nuclear RNP in yeast can be replaced with HARP or PRORP, a weaker growth phenotype was observed (15–17). Together these findings indicate the evolution of PRORPs is not due to advantages in enhanced activity or catalytic efficiency. Due to the presence of PRORPs in plants, metazoan mitochondria, and thermophilic prokaryotes, it is reasonable to hypothesize that PRORPs evolved to extend the range of environmental adaptive responses. The protein-only form of RNase P may offer protection from higher temperatures, alkaline environments, and oxidative stress. For example, *At* PRORP1, which localizes to the mitochondria, has been shown to have a 12 °C higher melting temperature than *At* PRORP2 and 3, which localize to the nuclei. *At* PRORP1 also maintains activity at temperatures up to 45 °C, while *At* PRORP2 and 3 lose all activity at 45 °C (18). This may correlate to studies that have suggested the mitochondria may be significantly warmer than the surrounding environment (19), potentially explaining why mitochondria rely on protein-only RNase P.

PRORPs also show a difference in substrate selectivity compared to their RNA-based counterparts, with almost all "weird" or "bizarre" non-canonical tRNAs seeming to be processed by PRORPs. Differences in substrate scopes may explain the various classes of protein-only RNase P enzymes. Metazoan mitochondrial tRNAs contain non-canonical structures, with differences such as the truncation or elimination of D and TψC loops and the absence of stabilizing tertiary interactions (20–23). To compensate for these non-canonical structures, metazoan (mt)tRNA interactors utilize altered modes of tRNA recognition. This has been shown for the aaRS, EF-Tu,

and metazoan mitoribosome, where it is believed co-evolution of interacting factors and (mt)tRNA led to altered modes of tRNA recognition. Metazoan (mt)RNase P is another interactor that co-evolved with metazoan (mt)tRNA, and similarly displays key differences in tRNA recognition to support the function of the non-canonical tRNA structures. Possibly this divergence in substrates and tRNA recognition strategy led to the retention of two forms of RNase P in metazoans. In addition, accessory proteins to protein-only RNase Ps may act as a chaperone, with multiple binding tRNAs either through a conformational selection between multiple accessible tRNA conformers or by assisting with its substrate target folding. The chaperones may also escort tRNAs to further processing enzymes or the translation complex or sequester tRNAs to prevent degradation. For example, the cryo-EM structure of the *Hs* (mt)RNase P complex demonstrates extensive contact between the proteins and the substrate, which may stabilize the pre-tRNA before maturation. Investigation into the PRORP interactome of *At* PRORP also reveals a network of interactions with other RNA processing proteins, further suggesting a role in escorting the tRNA substrates.

Discussion of the evolution and role of PRORPs is further complicated by the recent discovery of HARPs in *Aquifex aeolicus* and other thermophilic archaeal and bacterial organisms. Most organisms that encode HARP also encode RNA-based RNase P, with HARP knockout having less effects on cell growth than RNP knockdown. This raises questions about whether HARP is a redundant enzyme in these organisms or if it may have an additional function other than 5' tRNA processing. Currently available structural data shows HARPs have a highly conserved PIN-like metallonuclease domain 42 Å away from a positively charged putative binding region, but there is no structural data demonstrating interactions between tRNA and the active site. Why a completely distinct class of PRORPs evolved in thermophiles is unknown, although it has

been proposed that the ability of HARP to form a large homo-oligomeric complex may provide necessary thermal stability compared to the other classes of PRORP. Unfortunately, current research so far has not addressed the effects of temperature on the stability and activity of the proteins, as well as the structure and folding of the RNA substrates.

Although considerable progress has been made toward characterizing the atomic-level structure, kinetic activity, and substrate specificity of PRORPs, much remains to be discovered. The minimal PRORP discovered in *A. aeolicus* is now known to be not-so-minimal due to its ability to occupy large oligomeric states, but the conditions that cause these different oligomeric states have yet to be identified. New discoveries have demonstrated an increasingly complex interactome for PRORPs in *A. thaliana*, highlighting a need to understand protein-protein interactions between PRORPs and other tRNA-modifying and gene-regulating enzymes. Despite our growing understanding of the *A. thaliana* homologs of *Hs* (mt)RNase P, questions about the human PRORP linger. Although the molecular level impact of a handful of (mt)tRNA mutations on pre-tRNA maturation have been characterized, this accounts for only a fraction of known disease-related mutations in (mt)tRNAs and (mt)RNase P. Continuing the push to characterize these mutations will only strengthen our ability to develop ways to overcome them and treat mitochondrial disease.

4.4 Supplementary Information

SI Table 4.1: List of all currently available protein-only RNase P structures.

	PDB ID	Description	Reference
A. thaliana PRORPs	4G23	*A. thaliana* PRORP1	(Howard et al., 2012)
	4G24	*A. thaliana* PRORP1 with Mn	
	4G25	*A. thaliana* PRORP1 with SeMet and Sr	
	4G26	*A. thaliana* PRORP1 with Ca	
	5DIZ	*A. thaliana* PRORP2	(Karasik et al., 2016)
	5FT9	*A. thaliana* PRORP2	(Pinker et al., 2017)
	6LVR	*A. thaliana* PRORP1 PPR domain in complex with yeast tRNA^{Phe}	(Teramoto et al., 2020)
	6BV5	*A. thaliana* PRORP1 with Juglone	(Wu Meyers et al., 2022)
	6BV6	*A. thaliana* PRORP1 with Juglone	
	6BV8	*A. thaliana* PRORP1 with Juglone	
	6BV9	*A. thaliana* PRORP1 with Juglone	
Mitochondrial RNase P	4ROU	*H. sapiens* MRPP3	(F. Li et al., 2015)
	4XGL	*H. sapiens* MRPP3	(Reinhard et al., 2015)
	4XGM	*H. sapiens* MRPP3	
	7ONU	*H. sapiens* (mt)RNase P complex with (mt)pre-tRNA^{Tyr}	(Bhatta, et al. 2021)
HARPs	7F3E	*A. aeolicus* RNase P (Aq880) dodecamer	(Teramoto et al., 2021)
	7OG5	*H. halophila* RNase P (Hhal2243) dodecamer	(Feyh, et al. 2021)
	7E8J	*T. celer* RNase P dimer	(Y. Li et al., 2022)
	7E8K	*P. bacterium* RNase P dimer	
	7E8O	*P. bacterium* RNase P tetramer in complex with E. coli pre-tRNA^{His}	
	8SSF	*H. thermophilus* (Hth1307) RNase P tetramer	(Wilhelm et al., 2023)
	8SSG	*H. thermophilus* (Hth1307) RNase P tetradecamer	

4.5 References

1. Beychok, S. (1966) Circular Dichroism of Biological Macromolecules. *Science*. **154**, 1288–1299

2. Kahn, P. C. (1979) The interpretation of near-ultraviolet circular dichroism. in *Methods in Enzymology*, pp. 339–378, Enzyme Structure Part H, Academic Press, **61**, 339–378

3. Miles, A. J., Janes, R. W., and Wallace, B. A. (2021) Tools and methods for circular dichroism spectroscopy of proteins: a tutorial review. *Chem Soc Rev*. **50**, 8400–8413

4. Greenfield, N. J. (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc*. **1**, 2876–2890

5. Dunstone, M. A., and de Marco, A. (2017) Cryo-electron tomography: an ideal method to study membrane-associated proteins. *Philos Trans R Soc Lond B Biol Sci*. **372**, 20160210

6. Lučić, V., Leis, A., and Baumeister, W. (2008) Cryo-electron tomography of cells: connecting structure and function. *Histochem Cell Biol*. **130**, 185–196

7. Lučić, V., Rigort, A., and Baumeister, W. (2013) Cryo-electron tomography: The challenge of doing structural biology in situ. *Journal of Cell Biology*. **202**, 407–419

8. LeBlanc, S., Wilkins, H., Li, Z., Kaur, P., Wang, H., and Erie, D. A. (2017) Using Atomic Force Microscopy to Characterize the Conformational Properties of Proteins and Protein–DNA Complexes That Carry Out DNA Repair. *Methods Enzymol*. **592**, 187–212

9. Pleshakova, T. O., Bukharina, N. S., Archakov, A. I., and Ivanov, Y. D. (2018) Atomic Force Microscopy for Protein Detection and Their Physicochemical Characterization. *Int J Mol Sci*. **19**, 1142

10. Feyh, R., Waeber, N. B., Prinz, S., Giammarinaro, P. I., Bange, G., Hochberg, G., Hartmann, R. K., and Altegoer, F. (2021) Structure and mechanistic features of the prokaryotic minimal RNase P. *eLife*. **10**, e70160

11. Epple, S., El-Sagheer, A. H., and Brown, T. (2021) Artificial nucleic acid backbones and their applications in therapeutics, synthetic biology and biotechnology. *Emerg Top Life Sci*. **5**, 691–697

12. Liang, J. C., Bloom, R. J., and Smolke, C. D. (2011) Engineering biological systems with synthetic RNA molecules. *Mol Cell*. **43**, 915–926

13. Yi, P., Soper, T. J., and Woodson, S. A. (2012) RNase footprinting of protein binding sites on an mRNA target of small RNAs. *Methods Mol Biol*. **905**, 213–224

14. Howard, M. J., Liu, X., Lim, W. H., Klemm, B. P., Fierke, C. A., Koutmos, M., and Engelke, D. R. (2013) RNase P enzymes: divergent scaffolds for a conserved biological reaction. *RNA Biol*. **10**, 909–914

15. Taschner, A., Weber, C., Buzet, A., Hartmann, R. K., Hartig, A., and Rossmanith, W. (2012) Nuclear RNase P of *Trypanosoma brucei*: A Single Protein in Place of the Multicomponent RNA-Protein Complex. *Cell Reports*. **2**, 19–25

16. Weber, C., Hartig, A., Hartmann, R. K., and Rossmanith, W. (2014) Playing RNase P evolution: swapping the RNA catalyst for a protein reveals functional uniformity of highly divergent enzyme forms. *PLoS Genet*. **10**, e1004506

17. Nickel, A. I., Wäber, N. B., Gößringer, M., Lechner, M., Linne, U., Toth, U., Rossmanith, W., and Hartmann, R. K. (2017) Minimal and RNA-free RNase P in Aquifex aeolicus. *Proceedings of the National Academy of Sciences*. **114**, 11121–11126

18. Chen, T.-H., Sotomayor, M., and Gopalan, V. (2019) Biochemical Studies Provide Insights into the Necessity for Multiple *Arabidopsis thaliana* Protein-Only RNase P Isoenzymes. *Journal of Molecular Biology*. **431**, 615–624

19. Terzioglu, M., Veeroja, K., Montonen, T., Ihalainen, T. O., Salminen, T. S., Bénit, P., Rustin, P., Chang, Y.-T., Nagai, T., and Jacobs, H. T. (2023) Mitochondrial temperature homeostasis resists external metabolic stresses. *eLife*. 10.7554/eLife.89232

20. Helm, M., Brulé, H., Friede, D., Giegé, R., Pütz, D., and Florentz, C. (2000) Search for characteristic structural features of mammalian mitochondrial tRNAs. *RNA*. **6**, 1356–1379

21. Okimoto, R., and Wolstenholme, D. R. (1990) A set of tRNAs that lack either the TψC arm or the dihydrouridine arm: towards a minimal tRNA adaptor. *EMBO J*. **9**, 3405–3411

22. Suzuki, T., Nagao, A., and Suzuki, T. (2011) Human Mitochondrial tRNAs: Biogenesis, Function, Structural Aspects, and Diseases. *Annu. Rev. Genet.* **45**, 299–329

23. Watanabe, Y., Suematsu, T., and Ohtsuki, T. (2014) Losing the stem-loop structure from metazoan mitochondrial tRNAs and co-evolution of interacting factors. *Frontiers in Genetics*. 10.3389/fgene.2014.00109

Appendix A: Comparison of Predicted tRNA Secondary Structures and Energetics Across Six Species

Table A.1: Secondary structure and ΔG data of tRNAs from six species.[a]

Species names colored in orange or yellow are thermophiles. Species names colored in blue typically display optimal growth near 37 °C. Data under "Number of Nucleotides in the…" and "RNAFold" are colored with a gradient of red for above average, white for average, and green for below average.

tRNA Gene Information		Number of Nucleotides in the…[b]									RNAFold[c]
Species Name	GtRNAdb Gene Symbol + Link[d]	Acceptor Stem	Positions 8 & 9	D Stem	D Loop	Anticodon Stem	Anticodon Loop	Variable Loop	T Stem	T Loop	ΔG of Optimal 2⁰ Structure (kcal/mol)[e]
Average		6.94	2	3.69	8.94	4.98	7.03	7.72	4.96	7	-33.48
Median		7	2	4	8	5	7	5	5	7	-34
Mode		7	2	4	9	5	7	5	5	7	-33.3
Minimum		6	2	3	4	4	7	4	4	7	-57.4
Maximum		8	3	6	13	5	9	26	5	7	-20.2
A. aeolicus	tRNA-SeC-TCA-1-1	8	2	6	4	5	7	25	5	7	-57.4
H. thermophilus	tRNA-SeC-TCA-1-1	8	2	6	4	5	7	26	5	7	-55.1
A. aeolicus	tRNA-Ser-GCT-1-1	7	2	3	11	5	7	20	5	7	-53.9
T. Thermophilus	tRNA-Ser-CGA-1-1	7	2	3	11	5	7	20	5	7	-53.4
A. aeolicus	tRNA-Ser-TGA-1-1	7	2	3	11	5	7	21	5	7	-50
A. aeolicus	tRNA-Ser-GGA-1-1	7	2	3	11	5	7	20	5	7	-47.8
T. Thermophilus	tRNA-Ser-GCT-1-1	7	2	3	11	5	7	21	5	7	-47.7
A. aeolicus	tRNA-Ser-CGA-1-1	7	2	3	11	5	7	21	5	7	-47.5
T. Thermophilus	tRNA-Leu-CAA-1-1	7	2	3	11	5	7	15	5	7	-47

tRNA Gene Information		Number of Nucleotides in the...[b]									RNAFold[c]
Species Name	GtRNAdb Gene Symbol + Link[d]	Acceptor Stem	Positions 8 & 9	D Stem	D Loop	Anticodon Stem	Anticodon Loop	Variable Loop	T Stem	T Loop	ΔG of Optimal 2º Structure (kcal/mol)[e]
T. Thermophilus	tRNA-Leu-TAA-1-1	7	2	3	10	5	7	17	5	7	-47
E. coli	tRNA-SeC-TCA-1-1	8	2	6	4	5	7	22	5	7	-46.9
H. thermophilus	tRNA-Leu-CAA-1-1	7	2	3	11	5	7	15	5	7	-46.8
T. Thermophilus	tRNA-Ser-TGA-1-1	7	2	3	11	5	7	19	5	7	-46.8
A. aeolicus	tRNA-Leu-CAG-1-1	7	2	3	11	5	7	15	5	7	-46.7
A. aeolicus	tRNA-Leu-TAG-1-1	7	2	3	11	5	7	15	5	7	-46.7
A. aeolicus	tRNA-Tyr-GTA-1-1	7	2	3	11	5	7	16	5	7	-46.3
A. aeolicus	tRNA-Leu-GAG-1-1	7	2	3	11	5	7	16	5	7	-46.1
H. thermophilus	tRNA-Ser-GCT-1-1	7	2	3	11	5	7	21	5	7	-45.4
A. aeolicus	tRNA-Leu-CAA-1-1	7	2	3	11	5	7	16	5	7	-45.1
T. Thermophilus	tRNA-Leu-GAG-1-1	7	2	3	11	5	7	15	5	7	-44.4
T. Thermophilus	tRNA-Leu-TAG-1-1	7	2	3	11	5	7	13	5	7	-44.3
E. coli	tRNA-Ser-GCT-1-1	7	2	3	11	5	7	21	5	7	-43.9
T. Thermophilus	tRNA-Leu-CAG-1-1	7	2	3	11	5	7	15	5	7	-43.7
T. Thermophilus	tRNA-Ser-GGA-1-1	7	2	3	11	5	7	22	5	7	-43.5
H. thermophilus	tRNA-Leu-GAG-1-1	7	2	3	11	5	7	15	5	7	-43.2
H. thermophilus	tRNA-Ser-TGA-1-1	7	2	3	11	5	7	19	5	7	-43.2
A. aeolicus	tRNA-Leu-TAA-1-1	7	2	3	11	5	7	16	5	7	-43.1
H. thermophilus	tRNA-Ser-GGA-1-1	7	2	3	11	5	7	21	5	7	-42.9
E. coli	tRNA-Ser-CGA-1-1	7	2	3	11	5	7	18	5	7	-42.6
A. aeolicus	tRNA-Thr-TGT-1-1	7	2	4	8	5	7	5	5	7	-42.4
H. thermophilus	tRNA-Tyr-GTA-1-1	7	2	3	11	5	7	15	5	7	-42.2
H. thermophilus	tRNA-Leu-TAA-1-1	7	2	3	11	5	7	16	5	7	-42.1

tRNA Gene Information		Number of Nucleotides in the...[b]										RNAFold[c]
Species Name	GtRNAdb Gene Symbol + Link[d]	Acceptor Stem	Positions 8 & 9	D Stem	D Loop	Anticodon Stem	Anticodon Loop	Variable Loop	T Stem	T Loop		ΔG of Optimal 2° Structure (kcal/mol)[e]
B. subtilis	tRNA-Leu-TAA-1-1	7	2	3	11	5	7	17	5	7		-42
A. aeolicus	tRNA-Arg-CCG-1-1	7	2	4	9	5	7	5	5	7		-41.1
H. thermophilus	tRNA-Leu-CAG-1-1	7	2	3	11	5	7	15	5	7		-41
A. aeolicus	tRNA-Arg-TCT-1-1	7	2	4	9	5	7	5	5	7		-40
A. aeolicus	tRNA-Ile2-CAT-1-1	7	2	4	8	5	7	5	5	7		-39.4
T. Thermophilus	tRNA-Pro-TGG-1-1	7	2	4	9	5	7	5	5	7		-39.4
T. Thermophilus	tRNA-Arg-ACG-1-1	7	2	4	9	5	7	5	5	7		-39.3
H. thermophilus	tRNA-Ile2-CAT-1-1	7	2	4	7	5	7	5	5	7		-39.1
H. thermophilus	tRNA-Ser-CGA-1-1	7	2	3	11	5	7	19	5	7		-38.9
H. thermophilus	tRNA-Leu-TAG-1-1	7	2	3	11	5	7	14	5	7		-38.6
H. thermophilus	tRNA-Thr-TGT-1-1	7	2	4	8	5	7	5	5	7		-38.6
A. aeolicus	tRNA-Pro-TGG-1-1	7	2	4	8	5	7	5	5	7		-38.2
H. thermophilus	tRNA-Trp-CCA-1-1	7	2	4	8	5	7	5	5	7		-38
T. Thermophilus	tRNA-Val-TAC-1-1	7	2	4	9	5	7	5	5	7		-37.9
A. aeolicus	tRNA-Arg-ACG-1-1	7	2	4	9	5	7	5	5	7		-37.8
B. subtilis	tRNA-Ser-GCT-1-1	7	2	3	11	5	7	19	5	7		-37.8
A. aeolicus	tRNA-Pro-CGG-1-1	7	2	4	8	5	7	5	5	7		-37.7
H. thermophilus	tRNA-Arg-TCT-1-1	7	2	4	9	5	7	5	5	7		-37.6
H. thermophilus	tRNA-Phe-GAA-1-1	7	2	4	8	5	7	5	5	7		-37.3
A. aeolicus	tRNA-Phe-GAA-1-1	7	2	4	8	5	7	5	5	7		-37.3
A. aeolicus	tRNA-Trp-CCA-1-1	7	2	4	8	5	7	5	5	7		-37.3
E. coli	tRNA-Ser-GGA-1-1	7	2	3	11	5	7	16	5	7		-37.1
B. subtilis	tRNA-Leu-GAG-1-1	7	2	3	10	5	7	16	5	7		-36.8

tRNA Gene Information		Number of Nucleotides in the...[b]									RNAFold[c]
Species Name	GtRNAdb Gene Symbol + Link[d]	Acceptor Stem	Positions 8 & 9	D Stem	D Loop	Anticodon Stem	Anticodon Loop	Variable Loop	T Stem	T Loop	ΔG of Optimal 2º Structure (kcal/mol)[e]
B. subtilis	tRNA-Tyr-GTA-1-1	7	2	3	11	5	7	13	5	7	-36.8
A. aeolicus	tRNA-Cys-GCA-1-1	7	2	3	9	5	7	5	5	7	-36.6
A. aeolicus	tRNA-fMet-CAT-1-1	6	2	4	9	5	7	5	5	7	-36.6
E. coli	tRNA-Leu-TAA-1-1	7	2	3	11	5	7	15	5	7	-36.6
A. aeolicus	tRNA-Val-GAC-1-1	7	2	4	8	5	7	5	5	7	-36.5
E. coli	tRNA-Ile2-CAT-1-1	7	2	4	8	5	7	5	5	7	-36.5
A. aeolicus	tRNA-Val-TAC-1-1	7	2	4	8	5	7	5	5	7	-36.4
T. Thermophilus	tRNA-Tyr-GTA-1-1	7	2	3	11	5	7	14	5	7	-36.4
A. aeolicus	tRNA-Lys-CTT-1-1	7	2	4	8	5	7	5	5	7	-36.3
B. subtilis	tRNA-Ser-GGA-1-1	7	2	3	11	5	7	19	5	7	-36.3
T. Thermophilus	tRNA-Arg-CCT-1-1	7	2	4	9	5	7	5	5	7	-36.3
H. thermophilus	tRNA-Asn-GTT-1-1	7	2	4	8	5	7	5	5	7	-36.2
T. Thermophilus	tRNA-Gly-TCC-1-1	7	2	4	9	5	7	5	5	7	-36.2
H. thermophilus	tRNA-Gln-TTG-1-1	7	2	4	8	5	7	5	5	7	-36.1
A. aeolicus	tRNA-Glu-TTC-1-1	7	2	4	11	5	7	4	5	7	-36.1
E. coli	tRNA-Leu-TAG-1-1	7	2	3	11	5	7	13	5	7	-36.1
A. aeolicus	tRNA-Gly-TCC-1-1	7	2	4	8	5	7	5	5	7	-36
T. Thermophilus	tRNA-Glu-TTC-1-1	7	2	4	8	5	7	4	5	7	-36
E. coli	tRNA-Leu-GAG-1-1	7	2	3	11	5	7	15	5	7	-36
A. aeolicus	tRNA-Ile-GAT-1-1	7	2	4	9	5	7	5	5	7	-35.9
T. Thermophilus	tRNA-Glu-CTC-1-1	7	2	4	8	5	7	4	5	7	-35.8
H. thermophilus	tRNA-Val-TAC-1-1	7	2	4	8	5	7	5	5	7	-35.7
A. aeolicus	tRNA-Thr-CGT-1-1	7	2	4	7	5	7	5	5	7	-35.7

tRNA Gene Information		Number of Nucleotides in the...[b]									RNAFold[c]
Species Name	GtRNAdb Gene Symbol + Link[d]	Acceptor Stem	Positions 8 & 9	D Stem	D Loop	Anticodon Stem	Anticodon Loop	Variable Loop	T Stem	T Loop	ΔG of Optimal 2º Structure (kcal/mol)[e]
H. thermophilus	tRNA-Pro-TGG-1-1	7	2	4	8	5	7	5	5	7	-35.6
T. Thermophilus	tRNA-Pro-GGG-1-1	7	2	4	9	5	7	5	5	7	-35.6
T. Thermophilus	tRNA-Thr-GGT-1-1	7	2	4	8	5	7	5	5	7	-35.6
A. aeolicus	tRNA-Thr-GGT-1-1	7	2	4	8	5	7	5	5	7	-35.5
T. Thermophilus	tRNA-Met-CAT-1-1	7	2	4	9	5	7	5	5	7	-35.5
H. thermophilus	tRNA-Thr-GGT-1-1	7	2	4	8	5	7	5	5	7	-35.4
T. Thermophilus	tRNA-Arg-TCT-1-1	7	2	4	8	5	7	5	5	7	-35.4
A. aeolicus	tRNA-Lys-TTT-1-1	7	2	4	8	5	7	5	5	7	-35.3
T. Thermophilus	tRNA-Thr-TGT-1-1	7	2	4	8	5	7	5	5	7	-35.3
H. thermophilus	tRNA-Pro-CGG-1-1	7	2	4	8	5	7	5	5	7	-35.2
T. Thermophilus	tRNA-Ala-TGC-1-1	7	2	4	7	5	7	5	5	7	-35.2
T. Thermophilus	tRNA-Arg-CCG-1-1	7	2	4	9	5	7	5	5	7	-35.2
H. thermophilus	tRNA-Ile-GAT-1-1	7	2	4	9	5	7	5	5	7	-35
A. aeolicus	tRNA-Gly-GCC-1-1	7	2	4	7	5	7	5	5	7	-34.9
A. aeolicus	tRNA-His-GTG-1-1	7	2	4	8	5	7	5	5	7	-34.9
T. Thermophilus	tRNA-Gly-GCC-1-1	7	2	4	8	5	7	5	5	7	-34.9
H. thermophilus	tRNA-Gly-GCC-1-1	7	2	4	7	5	7	5	5	7	-34.8
H. thermophilus	tRNA-Arg-ACG-1-1	7	2	4	9	5	7	5	5	7	-34.7
H. thermophilus	tRNA-Thr-CGT-1-1	7	2	4	7	5	7	5	5	7	-34.7
A. aeolicus	tRNA-Gln-TTG-1-1	7	2	4	8	5	7	5	5	7	-34.6
T. Thermophilus	tRNA-Gly-CCC-1-1	7	2	4	8	5	7	5	5	7	-34.6
E. coli	tRNA-Val-TAC-1-1	7	2	4	8	5	7	5	5	7	-34.6
H. thermophilus	tRNA-Lys-CTT-1-1	7	2	4	8	5	7	5	5	7	-34.5

tRNA Gene Information		Number of Nucleotides in the...[b]									RNAFold[c]
Species Name	GtRNAdb Gene Symbol + Link[d]	Acceptor Stem	Positions 8 & 9	D Stem	D Loop	Anticodon Stem	Anticodon Loop	Variable Loop	T Stem	T Loop	ΔG of Optimal 2º Structure (kcal/mol)[e]
T. Thermophilus	tRNA-Ala-GGC-1-1	7	2	4	8	5	7	5	5	7	-34.5
T. Thermophilus	tRNA-Ile2-CAT-1-1	7	2	4	8	5	7	5	5	7	-34.5
S. cerevisiae	tRNA-Arg-ACG-1-1	7	2	3	10	5	7	5	5	7	-34.4
H. thermophilus	tRNA-Arg-CCG-1-1	7	2	4	9	5	7	5	5	7	-34.3
B. subtilis	tRNA-Ala-GGC-1-1	7	2	4	8	5	7	5	5	7	-34.2
H. thermophilus	tRNA-Arg-CCT-1-1	7	2	4	9	5	7	5	5	7	-34.1
A. aeolicus	tRNA-Pro-GGG-1-1	7	2	4	8	5	7	5	5	7	-34.1
E. coli	tRNA-Phe-GAA-1-1	7	2	4	8	5	7	5	5	7	-34.1
H. thermophilus	tRNA-Gly-TCC-1-1	7	2	4	7	5	7	5	5	7	-34
H. thermophilus	tRNA-Pro-GGG-1-1	7	2	4	8	5	7	5	5	7	-34
T. Thermophilus	tRNA-Thr-CGT-1-1	7	2	4	7	5	7	5	5	7	-34
T. Thermophilus	tRNA-Val-CAC-1-1	7	2	4	7	5	7	5	5	7	-34
A. aeolicus	tRNA-Arg-CCT-1-1	7	2	4	9	5	7	5	5	7	-33.9
A. aeolicus	tRNA-Asn-GTT-1-1	7	2	4	8	5	7	5	5	7	-33.9
T. Thermophilus	tRNA-Lys-CTT-1-1	7	2	4	8	5	7	5	5	7	-33.8
T. Thermophilus	tRNA-Pro-CGG-1-1	7	2	4	9	5	7	5	5	7	-33.8
H. thermophilus	tRNA-Gly-CCC-1-1	7	2	4	7	5	7	5	5	7	-33.7
H. thermophilus	tRNA-Val-GAC-1-1	7	2	4	7	5	7	5	5	7	-33.7
A. aeolicus	tRNA-Ala-TGC-1-1	7	2	4	8	5	7	5	5	7	-33.7
T. Thermophilus	tRNA-Ile-GAT-1-1	7	2	4	9	5	7	5	5	7	-33.6
E. coli	tRNA-Leu-CAG-1-1	6	2	3	11	5	7	15	5	7	-33.6
B. subtilis	tRNA-Ser-TGA-1-1	7	2	3	13	5	7	19	5	7	-33.5
A. aeolicus	tRNA-Met-CAT-1-1	7	2	4	9	5	7	5	5	7	-33.4

tRNA Gene Information		Number of Nucleotides in the...[b]									RNAFold[c]
Species Name	GtRNAdb Gene Symbol + Link[d]	Acceptor Stem	Positions 8 & 9	D Stem	D Loop	Anticodon Stem	Anticodon Loop	Variable Loop	T Stem	T Loop	ΔG of Optimal 2º Structure (kcal/mol)[e]
H. thermophilus	tRNA-Cys-GCA-1-1	7	2	3	9	5	7	5	5	7	-33.3
T. Thermophilus	tRNA-Val-GAC-1-1	7	2	4	9	5	7	5	5	7	-33.3
E. coli	tRNA-Ser-TGA-1-1	7	2	3	11	5	7	16	5	7	-33.3
B. subtilis	tRNA-Ile-GAT-1-1	7	2	4	9	5	7	5	5	7	-33.2
B. subtilis	tRNA-Leu-TAG-1-1	7	2	3	11	5	7	13	5	7	-33.1
T. Thermophilus	tRNA-Phe-GAA-1-1	7	2	4	8	5	7	5	5	7	-33.1
H. thermophilus	tRNA-fMet-CAT-1-1	6	2	4	9	5	7	5	5	7	-33
T. Thermophilus	tRNA-fMet-CAT-1-1	6	2	4	9	5	7	5	5	7	-33
B. subtilis	tRNA-Leu-CAA-1-1	7	2	3	11	5	7	13	5	7	-32.9
T. Thermophilus	tRNA-Ala-CGC-1-1	7	2	4	8	5	7	5	5	7	-32.9
T. Thermophilus	tRNA-Asn-GTT-1-1	7	2	4	8	5	7	5	5	7	-32.9
E. coli	tRNA-Glu-TTC-1-1	7	2	4	9	5	7	4	5	7	-32.8
A. Aeolicus	tRNA-Ala-GGC-1-1	7	2	4	8	5	7	5	5	7	-32.7
H. thermophilus	tRNA-Ala-GGC-1-1	7	2	4	8	5	7	5	5	7	-32.6
H. thermophilus	tRNA-Glu-TTC-1-1	7	2	4	11	5	7	5	5	7	-32.6
H. thermophilus	tRNA-Lys-TTT-1-1	7	2	4	9	5	7	5	5	7	-32.5
S. cerevisiae	tRNA-Leu-GAG-1-1	7	2	3	11	5	7	13	5	7	-32.4
T. Thermophilus	tRNA-Cys-GCA-1-1	7	2	3	9	5	7	5	5	7	-32.3
T. Thermophilus	tRNA-Lys-TTT-1-1	7	2	4	8	5	7	5	5	7	-32.3
S. cerevisiae	tRNA-iMet-CAT-1-1	7	2	4	7	5	7	5	5	7	-32.1
A. aeolicus	tRNA-Asp-GTC-1-1	7	2	4	10	5	7	5	5	7	-32
A. aeolicus	tRNA-Gly-CCC-1-1	7	2	4	7	5	7	5	4	7	-32
E. coli	tRNA-Arg-CCG-1-1	7	2	4	9	5	7	6	4	7	-32

tRNA Gene Information		Number of Nucleotides in the...[b]									RNAFold[c]
Species Name	GtRNAdb Gene Symbol + Link[d]	Acceptor Stem	Positions 8 & 9	D Stem	D Loop	Anticodon Stem	Anticodon Loop	Variable Loop	T Stem	T Loop	ΔG of Optimal 2º Structure (kcal/mol)[e]
H. thermophilus	tRNA-Ala-CGC-1-1	7	2	4	7	5	7	5	5	7	-31.9
H. thermophilus	tRNA-Asp-GTC-1-1	7	2	4	10	5	7	5	5	7	-31.9
T. Thermophilus	tRNA-Gln-TTG-1-1	7	2	4	7	5	7	4	5	7	-31.9
H. thermophilus	tRNA-His-GTG-1-1	7	2	4	8	5	7	5	5	7	-31.8
B. subtilis	tRNA-Leu-CAG-1-1	7	2	3	11	5	7	15	5	7	-31.8
B. subtilis	tRNA-Arg-CCG-1-1	7	2	4	9	5	7	7	5	7	-31.7
T. Thermophilus	tRNA-Trp-CCA-1-1	7	2	4	8	5	7	5	5	7	-31.7
E. coli	tRNA-Tyr-GTA-1-1	7	2	3	11	5	7	13	5	7	-31.7
T. Thermophilus	tRNA-Asp-GTC-1-1	7	2	4	9	5	7	5	5	7	-31.6
H. thermophilus	tRNA-Val-CAC-1-1	7	2	4	7	5	7	5	5	7	-31.5
B. subtilis	tRNA-fMet-CAT-1-1	6	2	4	9	5	7	5	5	7	-31.5
E. coli	tRNA-Pro-TGG-1-1	7	2	4	9	5	7	5	5	7	-31.4
E. coli	tRNA-Val-GAC-1-1	7	2	4	9	5	7	5	5	7	-31.4
E. coli	tRNA-Arg-TCT-1-1	7	2	4	8	5	7	5	5	7	-31.3
H. thermophilus	tRNA-Ala-TGC-1-1	7	2	4	7	5	7	5	5	7	-31.2
B. subtilis	tRNA-Asp-GTC-1-1	7	2	4	9	5	7	5	5	7	-31.2
E. coli	tRNA-Thr-GGT-1-1	7	2	4	8	5	7	5	5	7	-31.2
B. subtilis	tRNA-Ala-TGC-1-1	7	2	4	8	5	7	5	5	7	-31.1
E. coli	tRNA-Gly-GCC-1-1	7	2	4	8	5	7	5	5	7	-31
E. coli	tRNA-Leu-CAA-1-1	7	2	3	11	5	7	13	5	7	-31
B. subtilis	tRNA-Phe-GAA-1-1	7	2	4	8	5	7	5	5	7	-30.7
B. subtilis	tRNA-Val-GAC-1-1	7	2	4	8	5	7	5	5	7	-30.7
S. cerevisiae	tRNA-Gln-TTG-1-1	7	2	3	10	5	7	4	5	7	-30.6

tRNA Gene Information		Number of Nucleotides in the...[b]									RNAFold[c]
Species Name	GtRNAdb Gene Symbol + Link[d]	Acceptor Stem	Positions 8 & 9	D Stem	D Loop	Anticodon Stem	Anticodon Loop	Variable Loop	T Stem	T Loop	ΔG of Optimal 2º Structure (kcal/mol)[e]
S. cerevisiae	tRNA-Ser-AGA-1-1	7	2	3	10	5	7	14	5	7	-30.5
E. coli	tRNA-fMet-CAT-1-1	6	2	4	9	5	7	5	5	7	-30.3
E. coli	tRNA-Gly-CCC-1-1	7	2	4	7	5	7	4	5	7	-30.2
T. Thermophilus	tRNA-Gln-CTG-1-1	7	2	4	7	5	7	4	5	7	-29.9
H. thermophilus	tRNA-Met-CAT-1-1	7	2	4	9	5	7	5	5	7	-29.8
T. Thermophilus	tRNA-His-GTG-1-1	7	2	4	9	5	7	5	5	7	-29.8
S. cerevisiae	tRNA-Leu-CAA-1-1	7	2	3	11	5	8	13	5	7	-29.6
E. coli	tRNA-Ala-TGC-1-1	7	2	4	8	5	7	5	5	7	-29.6
S. cerevisiae	tRNA-Arg-CCG-1-1	7	2	4	8	4	7	4	5	7	-29.4
B. subtilis	tRNA-Glu-TTC-1-1	7	2	3	10	5	7	5	5	7	-29.3
E. coli	tRNA-Ile-GAT-1-1	7	2	4	9	5	7	5	5	7	-28.8
E. coli	tRNA-Pro-GGG-1-1	7	2	4	9	4	7	5	5	7	-28.8
B. subtilis	tRNA-Val-TAC-1-1	7	2	4	8	5	7	5	5	7	-28.7
S. cerevisiae	tRNA-Leu-TAA-1-1	7	2	3	11	5	7	15	5	7	-28.7
E. coli	tRNA-Ala-GGC-1-1	7	2	4	8	5	7	5	5	7	-28.7
S. cerevisiae	tRNA-Gly-GCC-1-1	7	2	4	7	5	7	4	5	7	-28.6
B. subtilis	tRNA-Met-CAT-1-1	7	2	4	9	5	7	5	5	7	-28.5
B. subtilis	tRNA-Pro-TGG-1-1	7	2	4	9	5	7	5	5	7	-28.5
E. coli	tRNA-Arg-ACG-1-1	7	2	4	9	5	7	5	5	7	-28.5
S. cerevisiae	tRNA-Arg-CCT-1-1	7	2	4	7	5	7	5	5	7	-28.4
E. coli	tRNA-Gly-TCC-1-1	7	2	3	10	5	7	4	5	7	-28.4
S. cerevisiae	tRNA-Ala-AGC-1-1	6	2	4	8	5	7	5	5	7	-28.3
S. cerevisiae	tRNA-Pro-TGG-1-1	6	2	3	9	5	8	5	5	7	-28.3

121

| tRNA Gene Information | | Number of Nucleotides in the...[b] | | | | | | | | | RNAFold[c] |
Species Name	GtRNAdb Gene Symbol + Link[d]	Acceptor Stem	Positions 8 & 9	D Stem	D Loop	Anticodon Stem	Anticodon Loop	Variable Loop	T Stem	T Loop	ΔG of Optimal 2º Structure (kcal/mol)[e]
E. coli	tRNA-Cys-GCA-1-1	7	2	3	9	5	7	5	5	7	-28.3
B. subtilis	tRNA-Arg-TCT-1-1	7	2	4	9	5	7	5	5	7	-28.2
S. cerevisiae	tRNA-Ser-CGA-1-1	7	2	3	10	5	8	14	4	7	-28.1
E. coli	tRNA-Asp-GTC-1-1	7	2	4	9	5	7	5	5	7	-28.1
E. coli	tRNA-Gln-TTG-1-1	7	2	3	9	5	7	5	5	7	-28.1
E. coli	tRNA-Lys-TTT-1-1	7	2	4	8	5	7	5	5	7	-28.1
B. subtilis	tRNA-Asn-GTT-1-1	6	2	4	7	5	7	5	4	7	-28
E. coli	tRNA-Gln-CTG-1-1	7	2	3	9	5	7	5	5	7	-28
E. coli	tRNA-Trp-CCA-1-1	7	2	4	8	5	7	5	5	7	-27.8
B. subtilis	tRNA-Ile2-CAT-1-1	7	2	4	9	5	7	5	5	7	-27.7
B. subtilis	tRNA-Arg-CCT-1-1	7	2	4	8	5	7	5	5	7	-27.6
S. cerevisiae	tRNA-Ile-AAT-1-1	7	2	3	11	5	7	5	5	7	-27.5
S. cerevisiae	tRNA-Gly-CCC-1-1	7	2	4	8	5	7	4	5	7	-27.4
E. coli	tRNA-Pro-CGG-1-1	7	2	4	9	5	7	5	5	7	-27.4
B. subtilis	tRNA-Arg-ACG-1-1	7	2	4	9	5	7	5	5	7	-27.3
S. cerevisiae	tRNA-Lys-CTT-1-1	7	2	4	8	5	7	5	5	7	-27.3
B. subtilis	tRNA-Cys-GCA-1-1	7	2	3	9	5	7	4	5	7	-27.1
B. subtilis	tRNA-Lys-TTT-1-1	7	2	4	8	5	7	5	5	7	-27.1
E. coli	tRNA-Arg-TCG-1-1	6	2	4	9	5	7	5	4	7	-27
B. subtilis	tRNA-Gly-GCC-1-1	7	2	4	7	5	7	5	5	7	-26.8
S. cerevisiae	tRNA-Asp-GTC-1-1	7	2	4	8	5	7	4	5	7	-26.6
S. cerevisiae	tRNA-Ser-TGA-1-1	7	2	3	10	5	7	14	4	7	-26.6
E. coli	tRNA-His-GTG-1-1	7	2	4	8	5	7	5	5	7	-26.6

tRNA Gene Information		Number of Nucleotides in the...[b]									RNAFold[c]
Species Name	GtRNAdb Gene Symbol + Link[d]	Acceptor Stem	Positions 8 & 9	D Stem	D Loop	Anticodon Stem	Anticodon Loop	Variable Loop	T Stem	T Loop	ΔG of Optimal 2º Structure (kcal/mol)[e]
E. coli	tRNA-Thr-CGT-1-1	7	2	4	8	4	7	6	5	7	-26.4
S. cerevisiae	tRNA-Ser-GCT-1-1	6	2	3	10	5	8	14	5	7	-26.3
S. cerevisiae	tRNA-Pro-AGG-1-1	6	2	3	9	5	7	5	5	7	-26.1
E. coli	tRNA-Arg-CCT-1-1	7	2	4	8	5	7	4	5	7	-26
B. subtilis	tRNA-Gln-TTG-1-1	7	2	3	9	5	7	5	5	7	-25.9
S. cerevisiae	tRNA-Thr-CGT-1-1	7	2	3	9	5	7	5	5	7	-25.5
B. subtilis	tRNA-Thr-TGT-1-1	7	2	4	8	5	7	5	5	7	-25.2
S. cerevisiae	tRNA-Gly-TCC-1-1	7	2	3	10	5	7	4	5	7	-25.2
S. cerevisiae	tRNA-Tyr-GTA-1-1	7	2	3	12	5	7	5	5	7	-25.1
B. subtilis	tRNA-Gly-TCC-1-1	7	2	4	7	5	7	4	4	7	-24.8
E. coli	tRNA-Met-CAT-1-1	7	2	4	9	5	7	5	4	7	-24.7
S. cerevisiae	tRNA-Cys-GCA-1-1	7	2	4	7	5	7	5	5	7	-24.6
E. coli	tRNA-Asn-GTT-1-1	7	2	4	8	5	7	5	5	7	-24.5
S. cerevisiae	tRNA-Ile-TAT-1-1	6	2	4	8	5	8	5	5	7	-24
S. cerevisiae	tRNA-Asn-GTT-1-1	7	2	3	11	5	7	5	5	7	-23.6
S. cerevisiae	tRNA-Leu-TAG-1-1	7	2	3	11	5	8	13	5	7	-23.6
B. subtilis	tRNA-Trp-CCA-1-1	7	2	4	7	5	7	4	5	7	-23.3
S. cerevisiae	tRNA-Arg-TCT-1-1	7	2	4	7	5	7	5	5	7	-23.3
B. subtilis	tRNA-His-GTG-1-1	7	2	3	10	5	7	5	5	7	-23.2
S. cerevisiae	tRNA-Ala-TGC-1-1	6	2	4	8	5	7	5	5	7	-23
S. cerevisiae	tRNA-Met-CAT-1-1	7	2	4	8	4	9	5	4	7	-23
E. coli	tRNA-Thr-TGT-1-1	7	2	4	8	5	7	5	4	7	-22.8
S. cerevisiae	tRNA-Trp-CCA-1-1	7	2	4	7	5	7	5	5	7	-22.7

tRNA Gene Information		Number of Nucleotides in the...[b]									RNAFold[c]
Species Name	GtRNAdb Gene Symbol + Link[d]	Acceptor Stem	Positions 8 & 9	D Stem	D Loop	Anticodon Stem	Anticodon Loop	Variable Loop	T Stem	T Loop	ΔG of Optimal 2° Structure (kcal/mol)[e]
S. cerevisiae	tRNA-Val-CAC-1-1	7	2	3	10	5	7	5	5	7	-22.6
S. cerevisiae	tRNA-Lys-TTT-1-1	7	2	4	8	5	7	5	5	7	-22.5
S. cerevisiae	tRNA-Phe-GAA-1-1	7	2	4	8	5	7	5	5	7	-22.4
S. cerevisiae	tRNA-Val-TAC-1-1	7	2	3	11	5	7	5	5	7	-22.2
S. cerevisiae	tRNA-His-GTG-1-1	7	2	4	8	5	7	4	5	7	-22.1
S. cerevisiae	tRNA-Thr-AGT-1-1	7	2	3	10	5	7	5	5	7	-21.5
S. cerevisiae	tRNA-Glu-TTC-1-1	7	2	3	10	5	7	4	5	7	-21.1
S. cerevisiae	tRNA-Val-AAC-1-1	7	2	3	11	5	7	5	5	7	-21.1
S. cerevisiae	tRNA-Thr-TGT-1-1	6	2	4	7	5	7	5	5	7	-20.4
S. cerevisiae	tRNA-Gln-CTG-1-1	7	2	3	10	5	7	4	4	7	-20.3
S. cerevisiae	tRNA-Glu-CTC-1-1	6	2	3	10	5	7	4	5	7	-20.2

[a] Values were obtained and compiled into a table by Matthew Yacoub and analyzed by Catherine Wilhelm.

[b] All secondary structure elements were colored with a color scale using green for the minimum value, white for the mode value, and red for the maximum value. The T loop had no variation.

[c] ΔG values were obtained using RNAFold (1).

[d] tRNA sequences linked here were obtained from GtRNAdb (2-4).

[e] ΔG values were sorted from lowest to highest and colored with a color scale from minimum value (green) to maximum value (red).

Figure A.1 (next page)**:** A stacked bar plot visualizing **Table A.1.**

This plot shows frequency of secondary structure elements larger than, equal to, or smaller than the mode in the six species detailed in **Table A.1.**
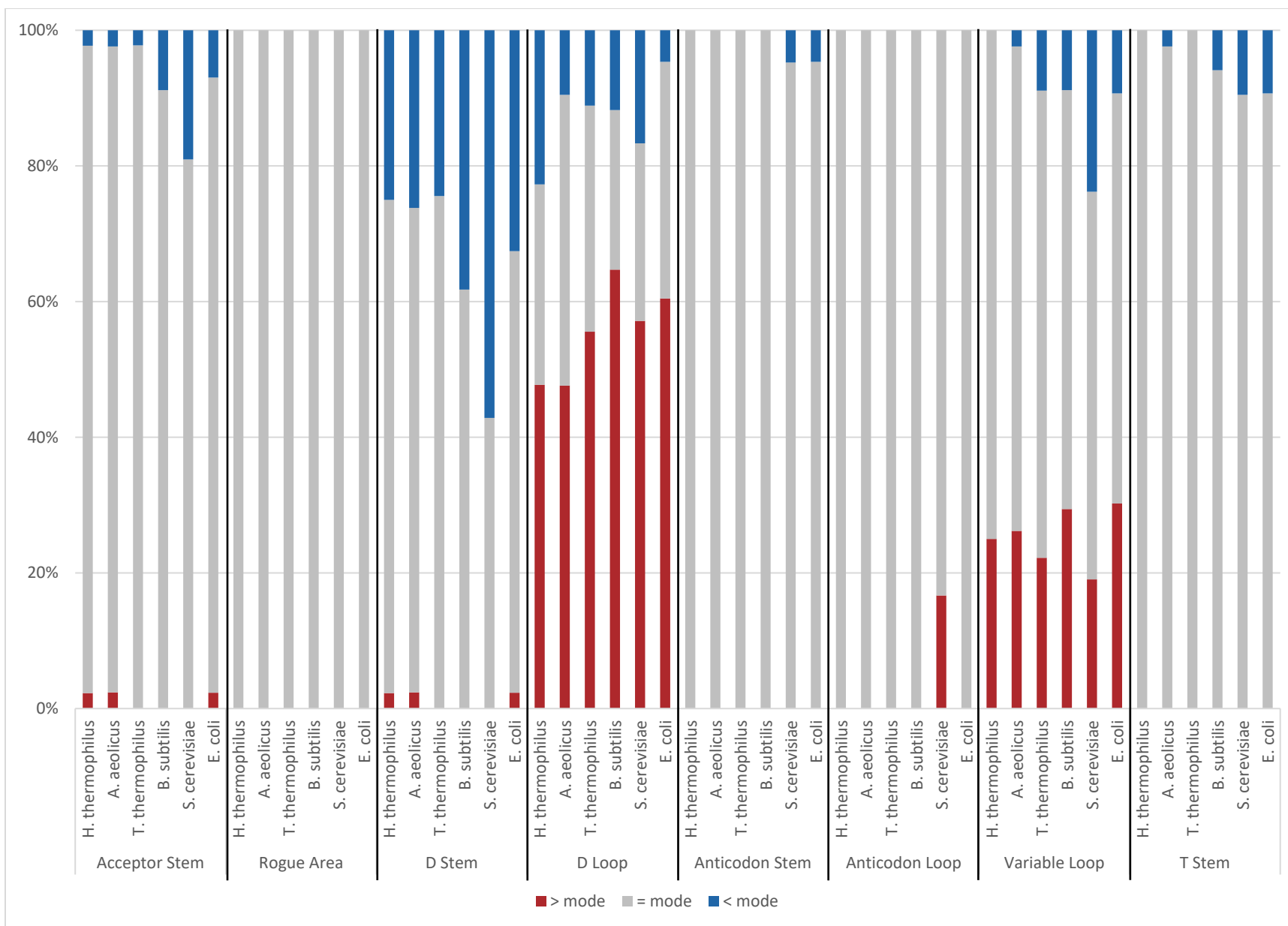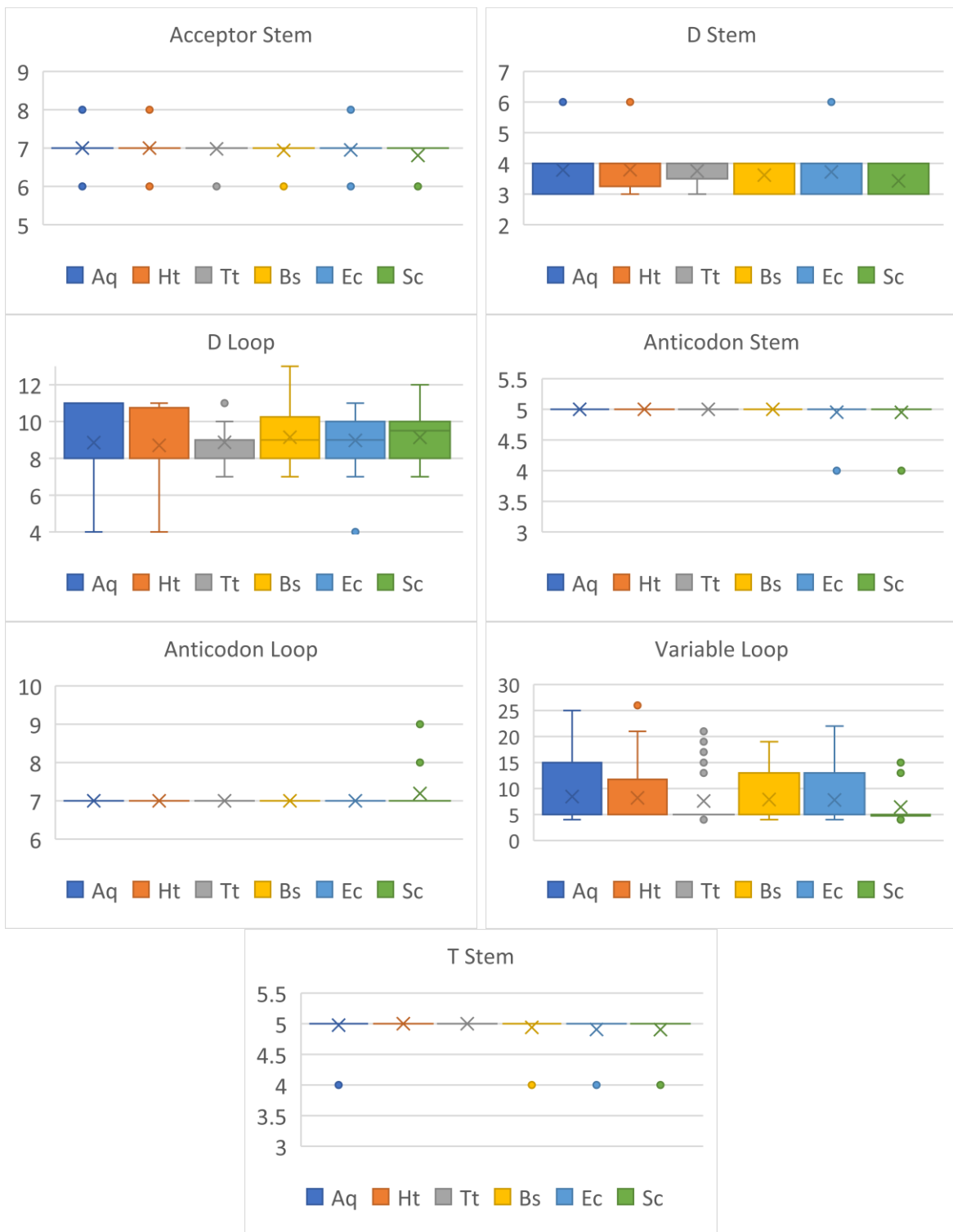
Figure A.2: Box and whisker plots visualizing **Table A.1**.

These plots show the frequency of secondary structure elements larger than, equal to, or smaller than the mode in the six species detailed in **Table A.1**.

References

1. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., and Hofacker, I. L. (2008) The Vienna RNA Websuite. *Nucleic Acids Res.* **36**, W70–W74

2. Chan, P. P., Lin, B. Y., Mak, A. J., and Lowe, T. M. (2021) tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077–9096

3. Lowe, T. M., and Chan, P. P. (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* **44**, W54–W57

4. Lowe, T. M., and Eddy, S. R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964