

Review

# The Role of 3' to 5' Reverse RNA Polymerization in tRNA Fidelity and Repair

Allan W. Chen <sup>1</sup>, Malithi I. Jayasinghe <sup>2</sup>, Christina Z. Chung <sup>1</sup>, Bhalchandra S. Rao <sup>2</sup>,  
Rosan Kenana <sup>1</sup>, Ilka U. Heinemann <sup>1,\*</sup> and Jane E. Jackman <sup>2,\*</sup> 

<sup>1</sup> Department of Biochemistry, The University of Western Ontario, 1151 Richmond Street, London, ON N6A 5C1, Canada; achen325@uwo.ca (A.W.C.); cchung88@uwo.ca (C.Z.C.); rkenana@uwo.ca (R.K.)

<sup>2</sup> Department of Chemistry and Biochemistry, Center for RNA Biology, Ohio State Biochemistry Program, The Ohio State University, Columbus, OH 43210, USA; jayasinghearachhige.1@osu.edu (M.I.J.); brao@arrakistx.com (B.S.R.)

\* Correspondence: ilka.heinemann@uwo.ca (I.U.H.); Jackman.14@osu.edu (J.E.J.)

Received: 28 February 2019; Accepted: 21 March 2019; Published: 26 March 2019

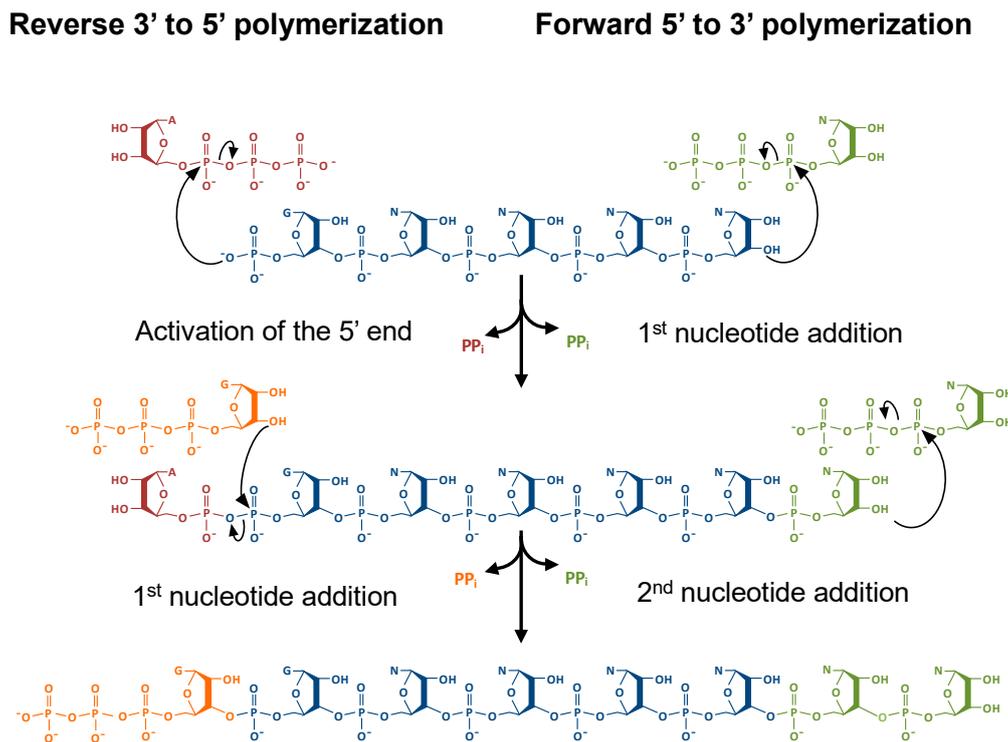


**Abstract:** The tRNA<sup>His</sup> guanylyltransferase (Thg1) superfamily includes enzymes that are found in all three domains of life that all share the common ability to catalyze the 3' to 5' synthesis of nucleic acids. This catalytic activity, which is the reverse of all other known DNA and RNA polymerases, makes this enzyme family a subject of biological and mechanistic interest. Previous biochemical, structural, and genetic investigations of multiple members of this family have revealed that Thg1 enzymes use the 3' to 5' chemistry for multiple reactions in biology. Here, we describe the current state of knowledge regarding the catalytic features and biological functions that have been so far associated with Thg1 and its homologs. Progress toward the exciting possibility of utilizing this unusual protein activity for applications in biotechnology is also discussed.

**Keywords:** reverse polymerization; tRNA editing; tRNA repair; protein engineering; synthetic biology

## 1. Introduction

During DNA strand synthesis, 5' to 3' polymerization conveys significant advantages, particularly in the removal of mismatched nucleotides by the exonuclease activity of DNA polymerases [1,2]. This 3' to 5' nucleotide removal regenerates a polymerizable 3'-OH end, whereas the removal of a 5' nucleotide generates a 5'-monophosphate, which is thought to be incapable of immediate replacement with another nucleotide (Figure 1) [2]. However, 3' to 5' or reverse polymerization is possible, despite the restrictions of extending a 5'-monophosphate end [3]. For such a 3'-5' polymerase to function, the 5'-monophosphate end could be activated by ATP, creating an intermediate that is competent for nucleophilic attack by the 3'-OH of an incoming nucleotide, which is then a mechanistically similar process to forward polymerization (Figure 1) [4,5]. Indeed, enzymes that are capable of catalyzing this reverse polymerase reaction, which are members of the tRNA<sup>His</sup> guanylyltransferase (Thg1) superfamily, have been recently discovered [6–10]. Despite the possible advantages of reverse polymerization, such as coupled leading strands in DNA replication as opposed to the generation of Okazaki fragments, 3' to 5' polymerization appeared to be more limited in biology, and was initially only observed to function in several reactions associated with tRNA repair and processing. However, newer data suggests broader substrate specificity exhibited by some members of the Thg1 family known as Thg1-like proteins (TLPs), thus raising the potential for additional roles for 3' to 5' polymerization yet to be discovered.



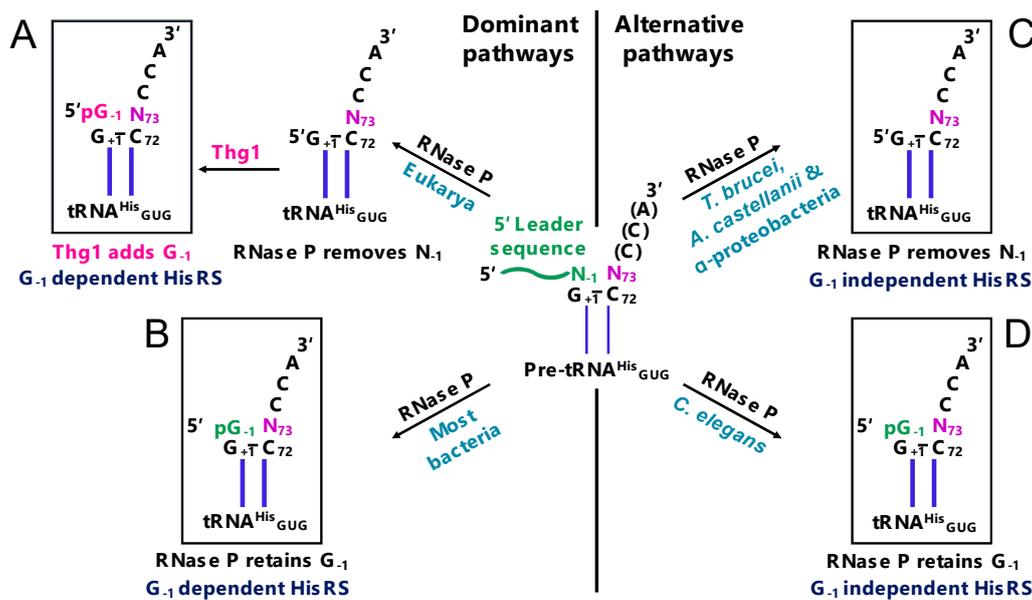
**Figure 1.** Forward 5' to 3' and reverse 3' to 5' polymerization are mechanistically similar. Ligation of a nucleotide to an RNA strand utilizes chemical energy that has been stored in the high-energy phosphoanhydride bonds in the nucleotide triphosphate. During forward polymerization, the triphosphate of the incoming nucleotide is hydrolyzed and provides energy to form a phosphodiester bond. For reverse polymerization, an initial activation step of a monophosphorylated RNA 5' end (such as by adenylation, which is shown here) is required; subsequent nucleotide additions utilize energy derived from hydrolysis of the phosphodiester bond of the RNA 5'-end triphosphate.

## 2. Thg1 Maintains tRNA<sup>His</sup> Aminoacylation Fidelity

### 2.1. $G_{-1}$ is a tRNA<sup>His</sup> Identity Element

Translational accuracy is dependent upon both tRNA selection by the ribosome and prior tRNA recognition and charging by a specific cognate aminoacyl-tRNA synthetase (aaRS) [11,12]. The side chain of the histidine residue, with its associated acid–base properties that contribute to both the protein structure and catalytic mechanisms, must be reliably and appropriately incorporated during translation [11,13]. Therefore, the aminoacylation of an amino acid to its respective tRNA must be specific, especially in the case of tRNA<sup>His</sup> due to its particular properties that are difficult to chemically mimic with another natural amino acid, possibly explaining why tRNA<sup>His</sup> relies on several key identity elements for proper aminoacylation.

For tRNA<sup>His</sup> in all domains of life, one of these identity elements is a distinctive extra guanine nucleotide at the 5'-end (the  $-1$  position) [14] (Figure 2). The 5' guanylate residue at the  $-1$  position ( $G_{-1}$ ) serves as a key identity element for histidyl-tRNA synthetase (HisRS), which charges tRNA<sup>His</sup> with its respective histidine [15–17]. In bacteria and many archaea, the  $G_{-1}$  residue is genomically encoded and transcribed in the precursor tRNA transcript, and subsequent cleavage of the 5' leader sequence by ribonuclease P (RNase P) yields a mature tRNA<sup>His</sup> with its identity-establishing  $G_{-1}$  element (tRNA<sup>His</sup> $G_{-1}$ ) (Figure 2) [14,18,19].



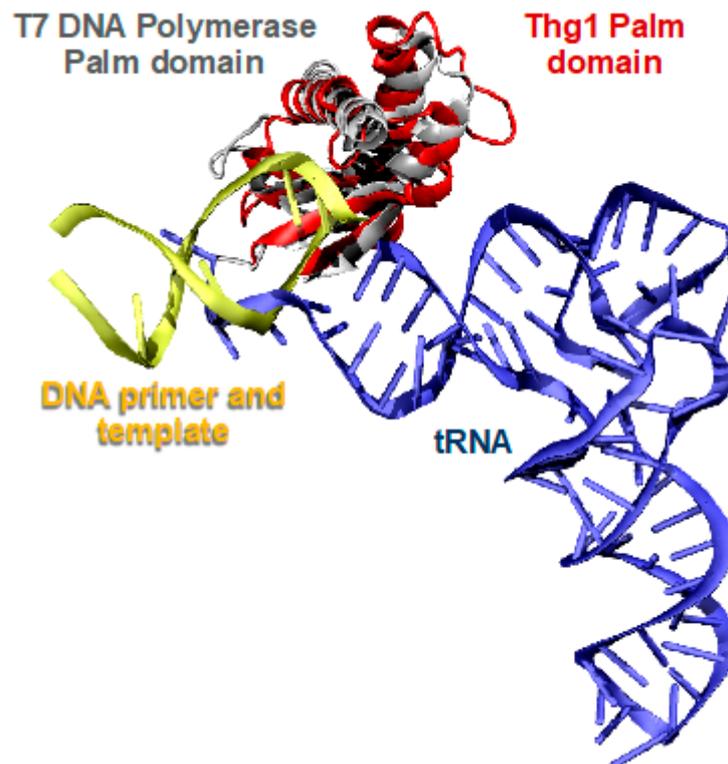
**Figure 2.** Different pathways to establish tRNA<sup>His</sup> identity. (A) tRNA<sup>His</sup> identity in many eukaryotes. RNase P removes N<sub>-1</sub> from pre-tRNA<sup>His</sup> during the removal of 5' leader sequence (shown in green). Then, tRNA<sup>His</sup> guanylyltransferase (Thg1) post-transcriptionally adds G<sub>-1</sub> (shown in pink). Histidyl-tRNA synthetase (HisRS) recognizes the Thg1-incorporated G<sub>-1</sub> for the accurate histidylation of tRNA<sup>His</sup> in eukaryotes. (B) tRNA<sup>His</sup> identity in bacteria. G<sub>-1</sub> is encoded in tRNA<sup>His</sup> genes in most bacteria, and RNase P retains the genomically encoded G<sub>-1</sub> during the removal of the 5' leader sequence. HisRS recognizes the RNase P-retained G<sub>-1</sub> for the accurate histidylation of tRNA<sup>His</sup> in bacteria. (C) tRNA<sup>His</sup> identity in several groups of α-proteobacteria, *Trypanosoma brucei* and *Acanthamoeba castellanii*. RNase P removes N<sub>-1</sub> from precursor tRNA<sup>His</sup> during the removal of the 5' leader sequence. An atypical HisRS in these organisms is capable of aminoacylating G<sub>-1</sub>-lacking tRNA<sup>His</sup> during histidylation. (D) tRNA<sup>His</sup> identity in *Caenorhabditis elegans* (an atypical eukaryote). G<sub>-1</sub> is encoded in tRNA<sup>His</sup> genes, and RNase P retains the genomically encoded G<sub>-1</sub>, which is similar to bacteria. HisRS in *C. elegans* is capable of aminoacylating both G<sub>-1</sub>-containing and G<sub>-1</sub>-lacking tRNA<sup>His</sup>.

In many eukaryotes, the establishment of tRNA<sup>His</sup> identity follows a different pathway. The G<sub>-1</sub> residue is not encoded in the tRNA<sup>His</sup> gene, but rather is added post-transcriptionally by Thg1 following RNase P cleavage (Figure 2A) [20,21]. Consistent with the necessity for the G<sub>-1</sub> determinant of tRNA<sup>His</sup> identity in protein fidelity, deletion or silencing of the gene encoding Thg1 results in severe growth defects in yeast, humans, and *Dictyostelium discoideum*, which is consistent with the wide conservation of Thg1 in Eukarya [21–23]. The addition of G<sub>-1</sub> by Thg1 in eukaryotes occurs opposite of the conserved terminal A<sub>73</sub> and is the result of a non-templated 3'-5' addition [20,21]. A few exceptions to the necessity of G<sub>-1</sub> have already been identified; a group of α-proteobacteria and several protozoan eukaryotes (*Acanthamoeba castellanii* and *Trypanosoma brucei*) do not genomically encode or post-transcriptionally add the G<sub>-1</sub> identity element; the absence of the otherwise highly conserved extra nucleotide is accommodated by an atypical HisRS recognition of tRNA<sup>His</sup> (Figure 2, right panel) [24–27].

## 2.2. Thg1 Structurally Resembles Canonical 5' to 3' Polymerases

Human Thg1 (HsThg1) is encoded by 269 amino acids with a calculated molecular weight of 32 kDa, but purification by gel exclusion chromatography eluted a protein of a molecular weight of ~165 kDa. This suggests the formation of a higher order multimer in solution, and is consistent with existing determinations of dimer-of-dimer tetrameric forms of active Thg1 enzymes [4,5,28–30]. Despite Thg1 enzymes sharing no obvious sequence similarity to known polymerases, a surprising structural homology exists between Thg1 and T7 DNA polymerase and DNA polymerase II (family

pol I and pol  $\alpha$ , respectively), as the enzymes share the same catalytic palm domain (Figure 3) [4,5]. The superposition of Thg1 with the aforementioned polymerases displays similar positioning of the three conserved carboxylate residues in the polymerase active site to the pol I family; HsThg1 carboxylates D29, D76, and E77 correspond to T7 DNA polymerases D475, D654, and E655. This similar carboxylate positioning suggests that the Thg1 mechanism of reverse polymerization may share features with the forward polymerization of the pol I family. Moreover, in the context of the overall structure, each monomeric subunit resembles a polymerase “hand” shape with the palm and fingers domains. This discovery revealed that forward and reverse polymerization can be accommodated by the same catalytic palm domain [4,5].



**Figure 3.** Forward and reverse polymerization is catalyzed by structurally similar enzymes but require opposing substrate orientation. The tRNA substrate (blue) of Thg1 (red) is in an opposing orientation compared to the DNA substrate and template (yellow) approaching the T7 DNA polymerase palm domain (grey).

Crystal structures of eukaryotic Thg1 enzymes showed that monomeric Thg1 is composed of a six-strand antiparallel  $\beta$ -sheet flanked by three or four  $\alpha$ -helices on each side, along with a protruding long arm composed of two antiparallel  $\beta$ -strands (Figure 3). Each monomer forms a dimer, mediated by salt bridge formation and hydrogen bonding between the  $\beta$ -sheet and an  $\alpha$ -helix [4,5]. Alanine scanning mutagenesis in *Saccharomyces cerevisiae* Thg1 (ScThg1) of conserved residues responsible for mediating dimer formation has shown strongly diminished  $G_{-1}$  addition activity [4,31]. Dimer-of-dimer formation stems from initial dimer formation; the  $\alpha$ -helix of the N-terminus in monomer A interacts with the nucleotide binding site of monomer B and ultimately forms an intertwined N-terminus segment, which again interacts with a symmetrical dimer to form its active tetrameric form [4,5]. The crystal structures of Thg1 from *Candida albicans* and *S. cerevisiae* corroborate similar findings in protein folding and quaternary structure, which is suggestive that Thg1 variants fold similarly, and are active in tetrameric form [4,5,28,30]

The cocrystal structure of *C. albicans* Thg1 (CaThg1) in complex with tRNA subsequently provided insight into the mechanistic details of reverse polymerization. Compared to forward

polymerases, such as the T7 DNA polymerase, the RNA substrate of Thg1 approaches the catalytic palm domain from the opposing site (Figure 3) [5]. Thus, the direction of polymerization is dependent on the direction of substrate approach to the catalytic core; Thg1 forces incoming nucleotides to approach from the opposite orientation of canonical forward polymerases, which is reflective of the overall domain organization of the enzyme (Figure 3). The fingers domain, which in part forms the nucleotide-binding site, is situated on opposite sides of the palm domain in Thg1, forcing a reversed substrate approach. Thus, the domain organization of Thg1 cannot accommodate forward polymerization, just as forward polymerases do not accommodate reverse polymerization [5]. Overall, these data revealed the molecular basis of reverse nucleotide addition, and showed that while reverse and forward polymerization can be catalyzed by the same catalytic core, substrate orientation may be the deciding factor in determining the directionality of nucleotide addition.

### 2.3. The Molecular Basis for tRNA Recognition

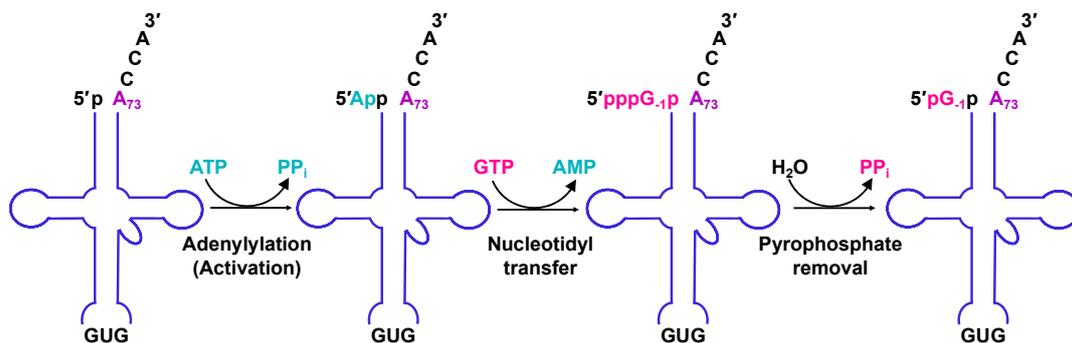
Thg1 recognizes tRNA<sup>His</sup> through its GUG anticodon, as demonstrated by the ability of Thg1 to add G<sub>-1</sub> to a mutagenized tRNA<sup>Phe</sup> that has been altered to contain the His anticodon (tRNA<sup>Phe</sup><sub>GUG</sub>) and subsequently validated through a cocrystal structure of *C. albicans* Thg1 in complex with tRNA<sup>Phe</sup><sub>GUG</sub>. The coordination of tRNA molecules occurs in a molar ratio of 4:2, where two tRNA molecules bind a Thg1 tetramer in parallel orientation, and each tRNA is coordinated by three subunits of the tetramer [5]. Common identity elements of the coordinated tRNA interact with the Thg1 tetramer; the acceptor stem is coordinated by the intertwined N-terminus of a Thg1 dimer, and the T arm is situated near polar residues on the rear surface of the catalytic core of the same dimer. This surface interaction in tRNA stabilization is analogous to the thumb domain in canonical forward polymerases. However, the anticodon, which is another distinctive identity element, is coordinated by the opposite dimer, and the G<sub>-1</sub> addition of this tRNA occurs solely in one dimer of the tetramer, specifically, the dimer that coordinates the acceptor stem and T arm positioning. The structural superposition of a crystal structure of ScThg1 and CaThg1-tRNA<sup>Phe</sup><sub>GUG</sub> discovered a conserved secondary structure in the fingers domain, and suggested conserved dual RNA-binding surfaces that were originally elucidated in CaThg1. Both the acceptor stem's sugar phosphate backbone and the anticodon loop are coordinated by the fingers domain; the latter fingers-anticodon interaction is base-specific [5].

Recognition of the tRNA<sup>His</sup> anticodon has been rationalized structurally through the CaThg1:tRNA<sup>Phe</sup><sub>GUG</sub> complex [5]. The G<sub>34</sub>, U<sub>35</sub>, and G<sub>36</sub> nucleotides that make up the tRNA<sup>His</sup> anticodon were all observed to be coordinated by specific residues of CaThg1. All three anticodon bases are tightly coordinated, and mutations of the coordinating amino acids lead to a disruption or reduction in enzyme activity [4,5,31]. The anticodon loop structure itself is stabilized by interactions with the sugar phosphate backbone of U<sub>35</sub>. Anticodon base G<sub>36</sub> is coordinated in a groove formed by two  $\alpha$ -helices flanking the central  $\beta$ -sheet and stabilized by a stacking interaction with H154, which is an essential residue in the eukaryotic-specific sequence motif HINNLYN [5,32]. The structure of Thg1 in complex with tRNA<sup>Phe</sup><sub>GUG</sub> suggests the molecular basis for the fairly stringent dependence of Thg1 on anticodon recognition, which is in agreement with its biological function of establishing tRNA<sup>His</sup> identity.

### 2.4. The Molecular Basis for Non-Watson-Crick G<sub>-1</sub> Addition: tRNA Activation

The resemblance between canonical forward polymerase structures and Thg1 in the overall structure raises questions about how exactly Thg1 performs reverse polymerization. In forward polymerization, the catalytic core coordinates the two catalytic metal ions with the first Mg<sup>2+</sup> (Mg<sup>2+</sup> A) promoting deprotonation of the 3'-OH in the polymerizing DNA strand, and the second Mg<sup>2+</sup> (Mg<sup>2+</sup> B) coordinating the triphosphate moiety of the incoming nucleotide. By analogy, Mg<sup>2+</sup> A and Mg<sup>2+</sup> B in Thg1 family polymerases use similar chemical features to catalyze the first step of G<sub>-1</sub> addition by activating the 5'-end of tRNA<sup>His</sup> with ATP (Figures 1 and 4). In this case, Mg<sup>2+</sup> A would promote the nucleophilic attack of the tRNA 5'-phosphate on the  $\alpha$ -phosphate of the activating ATP

nucleotide, which is also bound by  $Mg^{2+}$ . The exact mechanism of the activation step is not entirely known, but the GTP-bound crystal structures of HsThg1, ScThg1, and CaThg1 have elucidated binding interactions in the activation of the nucleotide-binding site and revealed a possible mechanism of how Thg1 differentiates between ATP and GTP in the initial activation of  $tRNA^{His}$  [4,5]. Nucleotidyl transfer in HsThg1 and ScThg1 is mediated largely by guanine base stacking against A37 and F42. The interaction is furthered by hydrogen bonding via amide, carbonyl, and the side chain moieties of D47, A43, and H34, respectively [4,5]. The superimposition of CaThg1–ATP and CaThg1–GTP cocrystal structures revealed that the nucleotide responsible for initial activation, ATP, resides deeper in the nucleotide-binding pocket than GTP, and interacts directly with D47 and K44. Mutational analysis of a ScThg1 D44A mutant greatly decreased the catalytic efficiency, suggesting that the D44 interaction is involved in  $G_{-1}$  addition activation [33].



**Figure 4.** Eukaryotic Thg1 catalyzes a three-step reaction. ScThg1 adds  $G_{-1}$  in three consecutive steps. First, the 5'-end of a monophosphorylated  $tRNA^{His}$  ( $p\text{-}tRNA^{His}$ ) is activated by an ATP to generate a 5'-adenylylated  $tRNA^{His}$  ( $App\text{-}tRNA^{His}$ ) intermediate (adenylation activation). Second, the 3'-hydroxyl of an incoming GTP nucleotide attacks the 5'-end of the  $App\text{-}tRNA^{His}$ , which then releases AMP and adds the GTP to yield a triphosphorylated  $tRNA^{His}$  ( $pppG_{-1}p\text{-}tRNA^{His}$ ) intermediate (nucleotidyl transfer). Third, pyrophosphate ( $PP_i$ ) from the  $pppG_{-1}p\text{-}tRNA^{His}$  is released, creating a mature  $pG_{-1}p\text{-}tRNA^{His}$  (pyrophosphate removal).

#### 2.5. The Molecular Basis for Non-Watson–Crick $G_{-1}$ Addition: Nucleotidyl Transfer

Thg1 family enzymes have a distinct site (at least partially separable from the activation site) for the binding of the incoming nucleotide used for the nucleotidyl transfer step (Figures 1 and 4). This can be clearly seen in HsThg1–dGTP and CaThg1–GTP crystal structures [4,5]. This observation is further supported by the kinetic data showing distinct functions for highly conserved residues in ScThg1. ScThg1 residues R27, K96, and R133 play a more significant role during the nucleotidyl transfer step based on more significant defects in the observed rate of this reaction with alanine variants compared to the other steps in  $G_{-1}$  addition [34]. However, these three residues only interact with the triphosphate moiety of the partially visualized incoming dGTP nucleotide in the HsThg1–dGTP crystal structure and the completely visualized GTP in the CaThg1–GTP structure. Due to the lack of direct contacts with the base or ribose of the GTP bound to the nucleotidyl transfer site in the CaThg1–GTP structure, neither of these structures explain how Thg1 specifically positions or recognizes its highly preferred nucleotide, GTP, over any other NTP, as observed repeatedly in many *in vitro* assays. Thus, the precise mechanism of Thg1-catalyzed nucleotidyl transfer to create the non-Watson–Crick  $G_{-1}\text{-}A_{73}$  base pair step is yet to be determined.

#### 2.6. The Molecular Basis for Non-Watson–Crick $G_{-1}$ Addition: Pyrophosphate Removal

In the final step of the reaction catalyzed during  $tRNA^{His}$  maturation by Thg1, the enzyme removes the 5'-pyrophosphate from the added  $G_{-1}$  nucleotide, yielding the 5'-monophosphate that is required for recognition by HisRS (Figure 4) [17,21,33]. This reaction requires the same two metal ion active sites as used for the previous two steps of the reaction, since alanine mutations to the metal

coordinating carboxylate residues completely eliminate the ability of the enzyme to catalyze this reaction in vitro [33]. However, additional residues that participate directly in the chemistry of this reaction have not yet been identified [33,35]. Interestingly, the removal of this pyrophosphate also removes an activated 5'-triphosphorylated end that could be extended further by the enzyme in the reverse polymerase reaction. Consistent with this idea, removal of the 5'-pyrophosphate by ScThg1 occurs much more efficiently in the context of the non-Watson–Crick base paired 5'-ends and 3'-ends of the tRNA than with a Watson–Crick base-paired end, thus preventing the further extension of this tRNA and limiting addition to the single essential G<sub>-1</sub> nucleotide [36].

### 2.7. Maintenance of tRNA<sup>His</sup> Fidelity

Maintaining tRNA identity is essential for translational fidelity, although a certain amount of mistranslation can be tolerated by the cell [12,37]. In eukaryotes, G<sub>-1</sub> addition is essential to establish tRNA<sup>His</sup> identity, and consequently, Thg1 is required for cell survival. In *S. cerevisiae*, the conditional depletion of Thg1 leads to the accumulation of unguanylated and uncharged tRNA<sup>His</sup> and growth arrest [8]. Interestingly, tRNA<sup>His</sup> in Thg1-depleted cells contains elevated levels of a 5-methylcytidine (m<sup>5</sup>C) modification at residues C<sub>48</sub> and C<sub>50</sub> [38]. While the change in nucleoside methylation is specific to tRNA<sup>His</sup>, it is most likely the result of the growth arrest rather than a direct consequence of G<sub>-1</sub> depletion. Upon conditional Thg1 depletion, these cells are arrested in prophase or G2. Thg1 was also shown to interact with the Orc2 component of the origin recognition complex, and loss of this interaction impairs DNA replication and nuclear division, although the precise molecular basis for this apparent connection of Thg1 to DNA metabolism remains unknown [39]. Thus, not surprisingly, Thg1 activity is essential in eukaryotes such as *S. cerevisiae*, which depend on Thg1 to maintain tRNA<sup>His</sup> identity.

## 3. Thg1-Like Proteins Function in tRNA Repair

### 3.1. Thg1-Like Proteins Are Functionally and Phylogenetically Distinct from Thg1

The biological function of Thg1 has been investigated in detail, but Thg1-like proteins (TLPs, which are also referred to as archaeal-type Thg1) only more recently gained attention. First identified in archaea, homologs have now been identified and characterized from several bacteria and eukaryotes. While related in sequence and structure, Thg1 and TLPs are phylogenetically and functionally distinct [32,40]. Where Thg1 is essential in many eukaryotes to establish tRNA<sup>His</sup> identity through the post-transcriptional addition of G<sub>-1</sub>, most bacteria and most archaea instead genomically encode the G<sub>-1</sub> identity element (Figure 2) [14,18–20,41]. Interestingly, in vitro analysis of bacterial and archaeal Thg1 homologs demonstrated that they are capable of catalyzing an analogous addition of G<sub>-1</sub> to bacterial/archaeal tRNA<sup>His</sup> transcripts that lack the G<sub>-1</sub> nucleotide, which would necessarily require the templated polymerization of G<sub>-1</sub> opposite to the C<sub>73</sub> that is universally found in these tRNA<sup>His</sup> [23,40]. Using these different biochemical capabilities and distinct sequence features as a functional classification system, Thg1 superfamily enzymes can thus be classified into two groups. The first group entails the bona fide Thg1 homologs, which are found only in Eukarya, and post-transcriptionally establishes tRNA<sup>His</sup> identity via G<sub>-1</sub>. The second group includes TLP homologs that are found in all three domains of life, and have not yet been implicated biologically in tRNA<sup>His</sup> identity, but for which this function would be in many cases redundant, if at all observed in vivo [23,32,40,42–44].

A phylogenetic analysis of candidate Thg1/TLP genes from all domains of life have grouped nearly all eukaryotic Thg1 variants into a single group; a second clade containing eukaryotic TLPs also contains archaeal and bacterial TLPs, which further supports the classification of Thg1 and TLPs as distinct enzymes [32,43,45]. A secondary trend further groups bacterial and archaeal Thg1 into two separate phylogenetic groups, and suggests that bacterial Thg1 was not descended from earlier bacterial ancestors [32,43,45].

### 3.2. The Discovery of TLPs in Bacteria and Archaea

The first characterization of a non-eukaryotic Thg1 homolog was carried out in *Methanosarcina acetivorans* [43]. Initially, the enzyme was annotated as two open reading frames split by a UAG Stop codon. While the enzyme activity of the split enzyme could be reconstituted in vitro, the protein is translated into a single protein in vivo, linking the two frames by a genetically encoded pyrrolysine at the UAG codon [43]. While in *M. acetivorans* the UAG codon could still be signaling for translation termination, it is unlikely that Thg1 is translated as two halves in vivo [46]. To date, Thg1 homologs belonging to the TLP clade have now been characterized biochemically from a diverse range of eight different bacterial and archaeal species. The common ability of all enzymes to catalyze strictly templated 3' to 5' addition reactions has been universally observed in assays of these enzymes with both tRNA<sup>His</sup> and non-tRNA<sup>His</sup> substrates [23,29,35,40,45,47]. The question of the precise nature of the biological substrates of these enzymes in their relevant organisms has not been entirely addressed, but some likely possibilities based on in vitro characterization so far include analogous tRNA<sup>His</sup> maturation reactions to those employed by eukaryotic Thg1 (albeit in many cases redundant with co-transcriptional mechanisms of obtaining the G<sub>-1</sub> nucleotide) and/or alternative RNA repair reactions catalyzed on substrates such as 5'-truncated tRNAs that remain to be identified.

The idea that bacterial and archaeal homologs of Thg1 might participate in G<sub>-1</sub> addition to tRNA<sup>His</sup> was an obvious extension of the known function of these enzymes in eukaryotes, and the initial in vitro characterization of several enzymes demonstrated activities that were consistent with this type of role [40,45,47]. It is notable that endogenous tRNA<sup>His</sup> genes in these species universally encoding a C<sub>73</sub> nucleotide in place of the A<sub>73</sub> found in eukaryotes was also consistent with the known preference of these enzymes to catalyze Watson–Crick-dependent 3' to 5' addition reactions. Thus, it was not surprising to see that TLPs were not able to efficiently add G<sub>-1</sub> to wild-type A<sub>73</sub>-containing eukaryotic tRNA<sup>His</sup> substrates, either in vitro or in vivo [23,40,45,47]. Interestingly, the *Bacillus thuringiensis* TLP homolog (BtTLP) is unusual compared to several other tested enzymes of this family, in that it is able to support the growth of a yeast  $\Delta thg1$  strain, presumably by acting in place of ScThg1 on tRNA<sup>His</sup> [23,40,47]. Indeed, sequences of tRNA<sup>His</sup> derived from the BtTLP-complemented *S. cerevisiae*  $\Delta thg1$  strain reveal that the tRNA contains almost exclusively U<sub>-1</sub>, instead of the canonical G<sub>-1</sub> residue [48]. In this case, the BtTLP evidently maintains its preference for incorporating Watson–Crick base-paired nucleotides by incorporating U<sub>-1</sub> across from the A<sub>73</sub> discriminator nucleotide. Although the minor growth defect of the BtTLP-expressing strains may then be attributed to the presence of the non-canonical U<sub>-1</sub> nucleotide on tRNA<sup>His</sup>, the ability of *S. cerevisiae* HisRS to accept other N<sub>-1</sub>-containing tRNA<sup>His</sup> substrates in vitro with only a fivefold to sixfold loss in efficiency is consistent with the observed viability of the BtTLP-expressing strain.

### 3.3. TLPs Catalyze Multiple Nucleotide Additions

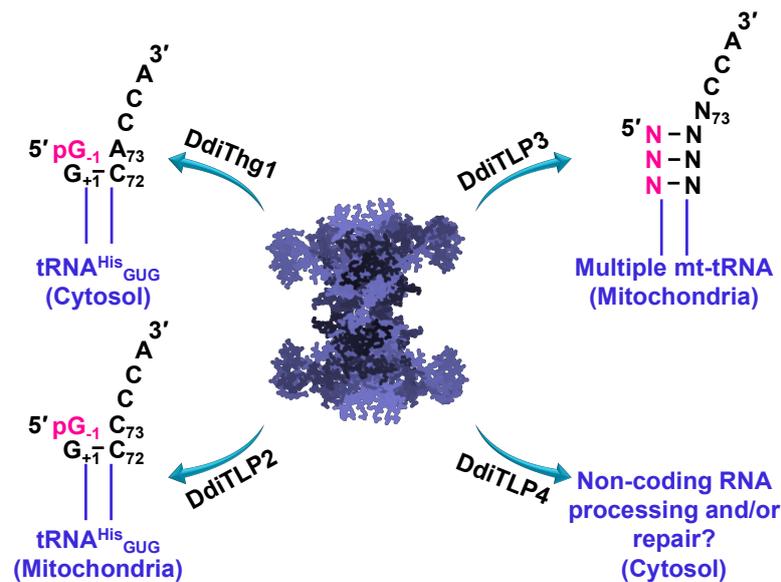
Interestingly, TLP homologs differ in terms of the number of nucleotide additions that are observed in vitro with tRNA transcripts mimicking the mature tRNA<sup>His</sup> that could be a substrate for this kind of activity in vivo. The ability to add multiple 5'-nucleotides to these types of tRNA<sup>His</sup> is a direct consequence of the presence of a C<sub>73</sub> discriminator, which results in the formation of three consecutive C-nucleotides in the 3'-C<sub>73</sub>CCA end. These types of substrates were initially shown to cause multiple G-addition reactions even in the context of eukaryotic Thg1, and several bacterial/archaeal homologs (the TLPs from *Methanopyrus kandleri*, *Pyrobaculum aerophilum*, and *B. thuringiensis*) behave similarly, catalyzing multiple G-additions [9,23,35,40,47]. However, other homologs (such as from *Myxococcus xanthus*, *Methanothermobacter thermoautotrophicus*, *Methanosarcina barkeri*, and *M. acetivorans*) appear to limit nucleotide addition to only a single guanine addition, despite the presence of an extended C-template at the tRNA 3'-end [40]. The molecular basis for these distinct behaviors is not yet known, but suggests that not all homologs in Bacteria and Archaea would react similarly in terms of any possible role in the maturation or maintenance of tRNA<sup>His</sup> [9,23,35,40,47].

The observation of alternative RNA substrate specificity exhibited by many enzymes in the TLP clade also suggested the possibility of non-tRNA<sup>His</sup> related functions for these enzymes. The identification of certain biochemical properties that distinguish TLPs from bona fide Thg1 counterparts lent support to these ideas. In addition to the previously described ability of TLPs to efficiently catalyze Watson–Crick base pair-dependent addition reactions, all of the bacterial and archaeal homologs investigated to date exhibit the preferential repair of 5'-truncated tRNAs instead of 5' nucleotide addition to full-length tRNA<sup>His</sup> [23,29,35]. This tRNA repair activity is not species-specific, as opposed to G<sub>-1</sub> addition in tRNA<sup>His</sup> identity establishment by bona fide Thg1 [23,26,31,42]. *M. acetivorans* TLP (MaTLP) has also demonstrated similar 5'-truncated tRNA repair as previously described with BtTLP; with preference to additions forming Watson–Crick base pairs [23,29,40]. Similarly, *Ignicoccus hospitalis* catalyzes an extended tRNA repair on truncated RNA substrates in vitro, adding up to 13 nucleotides in a templated reaction to restore a full tRNA<sup>His</sup> [35]. These data indicate that TLP function may indeed be found in tRNA repair rather than G<sub>-1</sub> addition.

#### 3.4. TLPs in Eukaryotes: Multiple TLPs Encoded by *Dictyostelium Discoideum*

Analysis of the eukaryotic slime mold, *Dictyostelium discoideum*, revealed that at least four Thg1 family enzymes appear to be present in the proteome (Figure 5). One of these genes is a bona fide Thg1 (DdiThg1), catalyzing the G<sub>-1</sub> addition to tRNA<sup>His</sup>, and thus establishing tRNA<sup>His</sup> identity [42,44]. The remaining three genes are characterized by sequence similarity and phylogeny as TLPs. Two of these TLPs (DdiTLP2 and DdiTLP3) are mitochondrial enzymes that catalyze distinct and non-redundant functions to add G<sub>-1</sub> to mitochondrial tRNA<sup>His</sup> (DdiTLP2) or to repair the 5'-end of mitochondrial tRNA (mt-tRNA) during a process known as tRNA 5'-editing (DdiTLP3) [44]. Both of these reactions take advantage of the 3' to 5' polymerase function that is the biochemically-preferred activity of TLP enzymes.

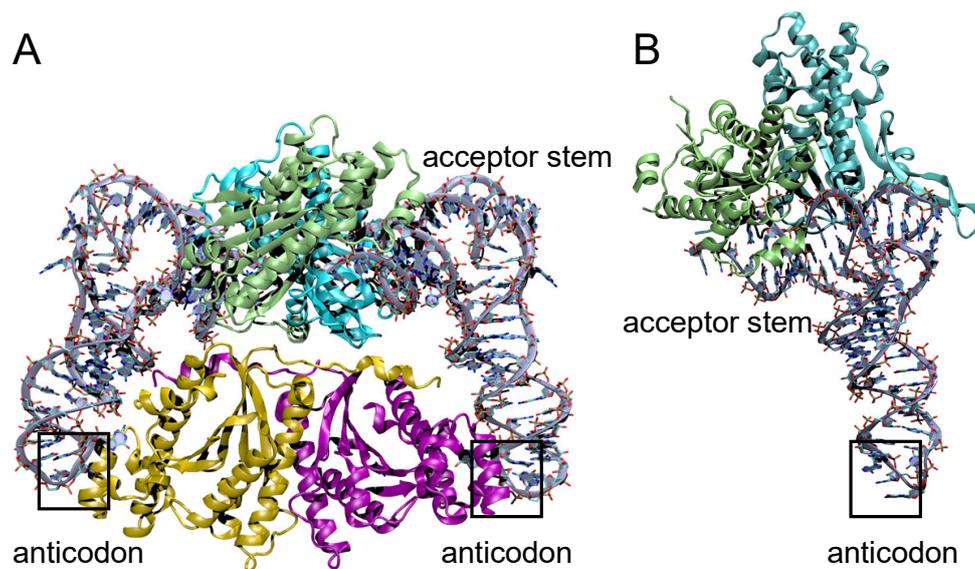
Although DdiTLP2 catalyzes a reaction that is on the surface very similar to that of a bona fide Thg1, there are several critical features that distinguish these two activities, which is consistent with the evolutionary distinct nature of these two enzymes. First, DdiTLP2 does not use the tRNA<sup>His</sup> GUG anticodon to recognize the tRNA for the addition of the G<sub>-1</sub> nucleotide, and its specific mechanism of tRNA<sup>His</sup> recognition remains unknown [44]. Second, the reaction catalyzed by DdiTLP2 is not essential for specifying tRNA<sup>His</sup> identity, since *D. discoideum*  $\Delta$ diti2 deletion strains are viable, but completely lack the G<sub>-1</sub> nucleotide on the mt-tRNA<sup>His</sup> [44]. DdiTLP3's role in tRNA 5'-editing also utilizes 3' to 5' polymerase function, but to repair mt-tRNA that have been truncated at their 5'-ends due to the removal of one or more incorrectly base-paired nucleotides encoded in the precursor tRNA [49]. This 5'-end repair step is essential for *D. discoideum*, and likely for many other single-celled eukaryotes that similarly encode mt-tRNA with 5'-mismatches [34,47,49–54]. Presumably, the TLPs encoded by these species are capable of participating in 5'-editing, although the identity of specific enzymes that participate in this process has so far only been demonstrated in *D. discoideum* [32,33,41–44]. Interestingly, the third TLP encoded in *D. discoideum* (DdiTLP4) catalyzes an essential function that remains unknown (Figure 5), although its ability to catalyze 3' to 5' polymerase activity on non-tRNA RNA substrates broadens the potential impact of these enzymes in terms of RNA processing and/or repair [44]. The continued functional characterization of TLP enzymes from diverse domains of life will be needed to provide a comprehensive picture of all of the biological reactions associated with this unusual family of enzymes (Figure 5).



**Figure 5.** Non-redundant physiological roles for Thg1 and Thg1-like proteins (TLPs) in *Dictyostelium discoideum* tRNA 5'-editing. DdiThg1 catalyzes cytosolic tRNA<sup>His</sup> maturation by incorporating G<sub>-1</sub> across from A<sub>73</sub>. DdiTLP2 catalyzes mitochondrial tRNA<sup>His</sup> maturation by incorporating G<sub>-1</sub> across from C<sub>73</sub>. DdiTLP3 catalyzes mitochondrial tRNA 5'-editing by repairing 5'-truncated tRNAs resulting from the removal of one or more mismatched nucleotides encoded in the precursor tRNAs. DdiTLP4 function remains unknown to date, but potentially this essential cytosolic enzyme is involved in non-coding RNA processing and/or repair in *D. discoideum*.

### 3.5. Structural Comparison of Thg1 and TLPs

The crystal structures of several members of the TLP clades of the Thg1 superfamily have been solved, and suggest many commonalities with Thg1, including an overall structural similarity to the *C. albicans* and *Homo sapiens* enzymes and the persistence of dimer-of-dimer quaternary structures [28, 29]. MaTLP assembles as a dimer-of-dimers similar to bona fide Thg1, yet uses a different mechanism of tRNA coordination. CaThg1 coordinates tRNA between both dimers of the tetramer, and binds Thg1–tRNA via 4:2 stoichiometry [54]. MaTLP independently binds one tRNA molecule per dimer, and does not seem to coordinate anticodon recognition by the opposing dimer, which is consistent with the notion of TLP repair activity (Figure 6) [23,40]. The present structures of the MaTLP and bona fide Thg1 variants differ most in tRNA coordination; the anticodon loop is not coordinated by the opposing dimer's fingers domain in TLPs [5,29]. MaTLP binds only the acceptor stem and T arm of its tRNA substrate; the flexible  $\beta$ -hairpin that coordinates the T arm has been speculated to enable the recognition of tRNA substrates with truncated acceptor stems [29]. The tRNA acceptor stem and T arm are coordinated by separate monomers in the dimer; the acceptor stem was hydrogen bonded to R136 and D137, and the triphosphate moiety at the 5'-end of the tRNA was bonded to the D21–K26 region and additionally coordinated by two Mg<sup>2+</sup>, which is similar to tRNA coordination by CaThg1 [5,29]. Expectedly, divalent cations of the MaTLP active site are coordinated by the carboxylates of D21 and D69, which is similar to previously reviewed bona fide Thg1 structures, other polymerases, and BfTLP. The T arm is recognized by the opposing monomer in the MaTLP dimer; phosphate groups of U<sub>55</sub> and G<sub>57</sub> are involved in hydrogen-bond interactions with the protruding long arm or  $\beta$ -hairpin region of MaTLP [29]. Computational analysis and model structures of MaTLP–tRNA complexes have shown truncated but not full-length tRNA molecules binding to MaTLP, suggesting that TLP molecules recognize 5'-truncated tRNA molecules via the flexible  $\beta$ -hairpin, and terminate elongation by recognizing the acceptor stem length of its substrates [29].



**Figure 6.** Structural comparison of Thg1 and TLPs. (A) The tetrameric eukaryotic Thg1 (monomers in green, teal, yellow and pink) coordinates two tRNAs molecules (grey) and binds both to the anticodon and the acceptor stem. (B) *Methanosarcina acetivorans* TLP forms a dimer (monomers in green and teal) coordinating on tRNA (grey). While the dimer binds the acceptor stem and T arm of its tRNA substrate, the anticodon loop is not coordinated by TLP enzymes.

The archaeon *I. hospitalis* encodes a minimized TLP homolog (IhTLP) with a molecular weight of 18 kDa, which is smaller than its archaeal cousin PaTLP (25 kDa) and the bona fide Thg1 from *C. alibicans* (35 kDa). IhTLP encodes the three conserved carboxylates that Thg1 and other polymerases require to coordinate two metal ions for catalysis, but shares little overall homology to other TLP enzymes. Furthermore, the fingers domain, which are responsible for the anticodon loop and acceptor stem coordination, is significantly minimized in IhTLP. IhTLP was found to be catalytically active in vitro, and catalyzes a significant tRNA repair reaction in vitro, adding up to 13 nucleotides to restore a truncated tRNA<sup>His</sup>, yet the in vivo function of IhTLP remains to be investigated [35]. While this mode of recognition is suitable for rationalizing TLP roles in repairing 5'-truncated tRNA, the prospect of alternative non-tRNA substrates for at least some members of this enzyme family raises additional questions about how this mechanism could be adapted more broadly to control 3' to 5' polymerase reactions.

### 3.6. 5'-End Activation and Nucleotidyl Transfer in TLPs

Many features of the overall mechanism of 3' to 5' nucleotide addition appear to be shared between Thg1 and TLP members of the enzyme family. Distinct residues that participate in 5'-end activation (*ScThg1*: K44, S75, N161) and nucleotidyl transfer (*ScThg1*: R27, K96, R133) are found in similar positions in the BfTLP structure, except for the N161 residue, which is absent in TLPs. The differing effects on distinct catalytic steps of the reaction that are observed upon the mutagenesis of these residues led to the notion of partially separable active sites for 5'-end activation and nucleotidyl transfer, which appears to also be the case for TLPs [4,5,28–30,33]. However, several molecular details distinguish these two types of enzymes. In terms of the first 5'-end activation step of the reaction, BfTLP and the TLP from *M. xanthus* (MxTLP) can utilize GTP to activate the 5'-end of tRNA<sup>His</sup> in vitro, whereas bona fide Thg1 enzymes are so far strictly ATP-dependent for catalyzing this reaction [6,23,28,40]. The crystal structures of TLP enzymes captured in the pre-activation conformation with either ATP or GTP nucleotides bound in the activation nucleotide-binding site have been determined, and confirm the roles for the residues implicated kinetically in the 5'-end activation reaction [28]. For the conserved lysine (K44 in *ScThg1*, K43 in BfTLP), this residue appears

to be important specifically for the activation reaction with ATP, and not used for the GTP activation reaction, since the alteration of this residue in the context of the bifunctional BtTLP only affected 5'-adenylation rates. Structural data support this ATP-specific role, since the side chain of K43 would predictably clash with the exocyclic 2'-amine of GTP bound in the active site. It is possible that ScThg1 K44 interacts with other yet-unidentified residues in the ScThg1 active site to ensure its ATP-dependent activity by preventing GTP binding in a catalytically competent conformation at this site.

Insight into the template-dependent nucleotidyl transfer step catalyzed by TLPs was provided by the structure of MaTLP bound to a GTP analog and a 5'-truncated triphosphorylated tRNA<sup>Phe</sup> (ppptRNA<sup>Phe</sup> $\Delta_1$ ), which mimics the activated tRNA ready for the nucleotidyl transfer. The incoming nucleotide uses Watson–Crick base-pairing and base-stacking interactions to facilitate its incorporation to the tRNA substrate [28]. In addition, binding of the incoming nucleotide seems to promote a significant structural change in the 5'-end of the tRNA substrate to accommodate the nucleotide incorporation across from the templating nucleotide. Also, no interactions between the enzyme and the base moiety of the GTP analog were observed here, supporting the idea of template-dependency exhibited by TLPs for nucleotide addition to their substrates, which is presumably distinct from the expected base-recognition capabilities for Thg1 homologs to catalyze template-independent GTP addition. In TLPs, coordination of the Mg<sup>2+</sup> A to the 3'-OH of incoming NTP facilitates the nucleophilic attack on the  $\alpha$ -phosphate of the activated tRNA substrate, which is analogous to the role of this residue in canonical polymerases. Meanwhile, the triphosphate of the incoming NTP coordinates to a third metal ion (Mg<sup>2+</sup> C). It is arguable that the Mg<sup>2+</sup> C seen in the active site of Thg1/TLP structures stabilizes the incoming NTP in the absence of metal coordination from other two Mg<sup>2+</sup> ions (Mg<sup>2+</sup> A and Mg<sup>2+</sup> B), which are already coordinated to either the activating ATP that resides in the activation site or the activated tRNA substrate. Appearance of a third metal ion has also been observed in DNA pol  $\beta$  and DNA pol  $\eta$  [55,56]. Although these polymerases use the canonical two-metal ion mechanism, the third metal ion has been proposed to facilitate phosphoryl transfer to the nucleophilic 3'-OH group of the growing polynucleotide chain. Similarly, the use of a third Mg<sup>2+</sup> ion (Mg<sup>2+</sup> C) in TLPs seems to facilitate nucleophilic attack on the  $\alpha$ -phosphate of the activated tRNA to promote PP<sub>i</sub> release upon subsequent nucleotidyl transfer.

#### 4. Synthetic Biology Applications and Thg1/TLP Engineering

How TLP enzymes distinguish between truncated tRNA and full-length tRNA in the context of 5'-tRNA elongation or repair and how the elongation reaction is properly terminated in all cases remains to be fully solved. The proper termination of Thg1 enzymes after adding only the single G nucleotide to tRNA<sup>His</sup> is intimately connected with the acquisition of the A<sub>73</sub> discriminator nucleotide of tRNA<sup>His</sup> in eukaryotes, as opposed to the universally conserved (and likely ancestral) C<sub>73</sub> that is found in Bacteria and Archaea. Indeed, the molecular basis for the ability of the G<sub>-1</sub>:A<sub>73</sub> base pair to trigger the termination of the addition in ScThg1 is demonstrated to result from ScThg1's highly efficient removal of the 5'-triphosphorylated end from tRNA species that terminate in this non-Watson–Crick base pair, as described above [36]. This mechanism for termination is unlikely to be conserved among TLPs, since these enzymes strictly add Watson–Crick base pairs that are not efficient substrates for the 5'-pyrophosphatase activity [36]. The previously demonstrated capability of TLPs to catalyze extended reverse RNA polymerization [23,29,35,40] makes these proteins promising candidates for protein engineering to harness their ability to extend RNAs in the 3' to 5' direction in a template-dependent manner.

##### 4.1. TLPs Exhibit Broader RNA Recognition Properties Than Thg1 Homologs

The ability to act on a broader range of tRNA substrates than the tRNA<sup>His</sup>-specific Thg1 homologs naturally positions TLP enzymes as candidates for engineering 3' to 5' polymerases for diverse functions. Understanding the basis for RNA recognition by various members of the Thg1 superfamily is a key element to these efforts. While recognition of the anticodon GUG is a major determinant for

Thg1-catalyzed  $G_{-1}$  addition to tRNA<sup>His</sup>, the TLPs that have been tested to date do not depend on the presence of this sequence to be used as substrates for 3' to 5' polymerization [10,23]. This ability to act independently of the specific tRNA<sup>His</sup> anticodon nucleotides is a necessary feature for TLPs to act outside of tRNA<sup>His</sup> maturation, as has already been demonstrated by the physiological function of DdiTLP3 in mt-tRNA 5'-editing [42,44]. For this 5'-editing reaction, the ability to act on any tRNA is required, since many different tRNA species contain the genomically encoded 5'-mismatches that must be removed by 5'-editing and then repaired by the TLP prior to participation in translation [34,49,51–53]. Indeed, the pattern of tRNA species that have been demonstrated to undergo editing in the many diverse eukaryotes where this process has been investigated so far is quite different from organism to organism, and can involve from only a few species to nearly all of the mt-tRNA, thus suggesting that flexibility in tRNA recognition is important for this function [50]. The idea that some TLPs may act on other non-tRNA substrates also raises new questions for these enzymes and their ability to recognize various RNAs. Thus, TLP activity is not restricted to tRNA<sup>His</sup> substrates, and may be engineered to further broaden their substrate specificity.

#### 4.2. Steps Toward Utilizing TLPs for Targeted 3'-5' Addition Reactions

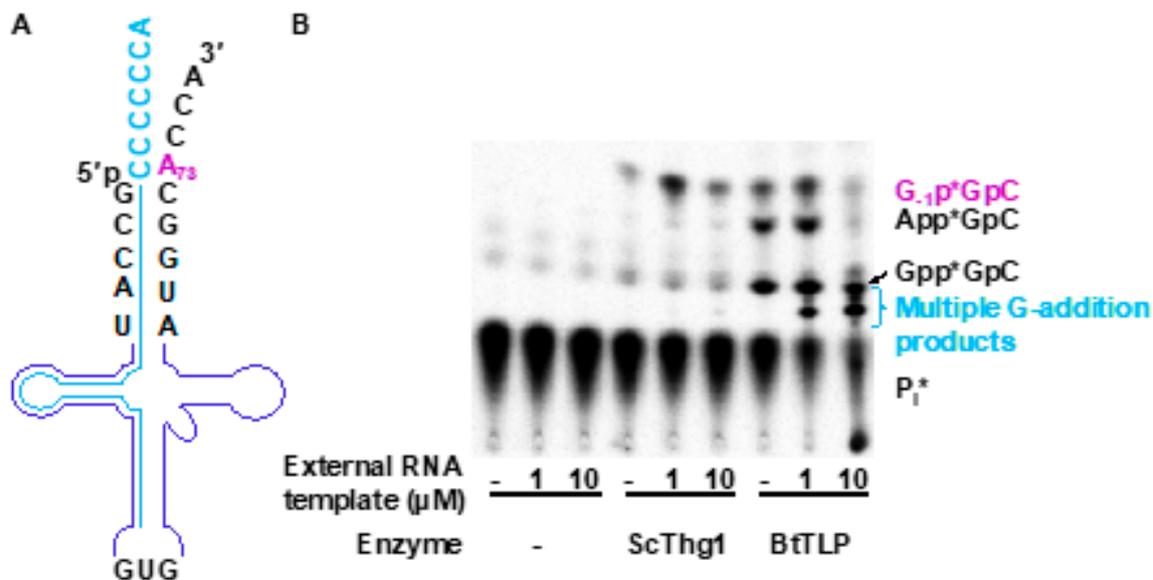
Although eukaryotic Thg1 is mostly restricted to acting on tRNA<sup>His</sup>, bacterial and archaeal TLPs have been shown to exhibit broader RNA substrate specificity, suggesting that the 3' to 5' polymerase activity of TLPs could potentially be directed to site-specifically label diverse RNA substrates at their 5'-ends. TLPs are capable of adding all four nucleotides in a template-dependent fashion [23,29,35,40], which makes them true reverse polymerases. This could in principle also be extended to add labeled or modified nucleotides at specific positions of a given RNA substrate. This approach has previously been successful in 3'-biotinylation by the terminal nucleotidyltransferase Cid1 [38]. Other applications that we can envision include reverse polymerization on additional substrates, which would allow for applications such as the site-specific incorporation of nucleotides to the 5'-ends of tRNA transcripts that are difficult to incorporate during in vitro synthesis due to the limitations of the commonly used T7 RNA polymerase. Here, we describe several approaches that have been used successfully for this type of application.

One critical step in Thg1/TLP engineering is to broaden substrate specificity toward RNAs other than tRNA. While the tRNA<sup>His</sup> GUG anticodon is not required for TLP activity, recognition of the overall tertiary tRNA structure is a conserved element for Thg1 and TLPs. Desai et al. developed a split tRNA approach, which was later successfully implemented by Nakamura et al., dividing the tRNA through the D-loop. Here, the tRNA structure is mostly provided by a guide RNA, complementary to the RNA of interest to be guanylated [35,54]. The guided RNA approach has been successful to lead RNaseP to mRNA substrates and cleave pathogenic RNAs in cancer [57]. Thus, this approach could be used to direct Thg1 to any given RNA 5'-end when provided with a complementary sequence, resulting in a tRNA-like structure.

An alternative approach provides an external RNA-guide template, which is used to anneal to the full-length tRNA, ideally disrupting the structure in the aminoacyl acceptor and D-stems, and thus providing a template to add the desired nucleotide(s) to the tRNA. To demonstrate the feasibility of this approach, we exploited the known inability of TLP enzymes to efficiently add a non-Watson–Crick base-paired  $G_{-1}$  to wild-type tRNA<sup>His</sup> with an A<sub>73</sub> discriminator nucleotide [23,47]. Then, by adding an additional RNA oligonucleotide that is complementary to the tRNA sequence, we provided an alternative template that could possibly direct the addition of a nucleotide(s) to the tRNA 5'-end in the Watson–Crick-dependent manner that is preferred by TLPs. We reasoned that if TLP enzymes are capable of accommodating an external oligonucleotide template, 3'-5' polymerase activity could potentially be directed to add nucleotides to the 5'-end of any given RNA substrate [49].

To test this, we used labeled in vitro transcribed *S. cerevisiae* tRNA<sup>His</sup> (Sc-tRNA<sup>His</sup>) (Figure 7A). In the absence of any added oligonucleotide, ScThg1 adds  $G_{-1}$ , forming a non-Watson–Crick base pair, as evident from the  $G_{-1}p^*GpC$  product that is observed after G-addition (Figure 7B). BtTLP

only weakly adds  $G_{-1}$ ; however, as observed previously, it instead accumulates 5'-adenylylated and 5'-guanylylated intermediates that are the products of the first activation reaction (Figure 7B). Then, varied concentrations (1  $\mu\text{M}$  and 10  $\mu\text{M}$ ) of an externally provided 38-mer oligonucleotide template that is complementary to 31 nucleotides on the 5'-end of Sc-tRNA<sup>His</sup> were added to the assays. When this oligonucleotide anneals to Sc-tRNA<sup>His</sup>, it will create a 7-nucleotide 3'-overhang that can serve as a template for nucleotide addition to the 5'-end of the Sc-tRNA<sup>His</sup> (Figure 7A). Here, the 3'-CCCCCA overhang has been chosen due to the preferential G–C base pairing activity exhibited by TLP enzymes.



**Figure 7.** An external guide RNA oligonucleotide-mediated approach to manipulate nucleotide addition by *Bacillus thuringiensis* TLP homolog (BtTLP). (A) Schematic representation of tRNA<sup>His</sup> (backbone in purple) hybridized to the external guide RNA oligonucleotide template (shown in blue). (B) Activity assay using the external guide RNA template and analyzed by thin-layer chromatography (TLC). Identities of the labeled reaction products are consistent with the known migration patterns of these species that have been validated by previous assays [23]. Nucleotide addition was tested using *Saccharomyces cerevisiae* Thg1 (ScThg1) and BtTLP with varying concentrations of the external guide RNA template (0  $\mu\text{M}$ , 1  $\mu\text{M}$ , and 10  $\mu\text{M}$ , as indicated). In the absence of the guide RNA template, ScThg1 efficiently adds  $G_{-1}$  and in contrast, BtTLP only weakly adds  $G_{-1}$  across from A<sub>73</sub> (evident from the amount of the product  $G_{-1}\text{p}^*\text{GpC}$ , which is labeled in purple). BtTLP instead accumulates 5'-adenylylated and 5'-guanylylated intermediates (evident from the products App\*GpC and Gpp\*GpC, which are labeled in black respectively). In the presence of the external guide RNA template, ScThg1 still maintains its ability to add a single G nucleotide across from A<sub>73</sub> in *S. cerevisiae* tRNA<sup>His</sup> (Sc-tRNA<sup>His</sup>, as evident from the product  $G_{-1}\text{p}^*\text{GpC}$  shown in purple), whereas BtTLP adds multiple G nucleotides using the 3'-CCCCCA of the external guide RNA oligonucleotide as a template. This multiple addition is evident from the appearance of several lower migrating products, which are labeled in blue, and are absent in ScThg1-containing reactions. The precise number of G-residues added to the tRNA cannot be determined due to the inability of this TLC system to resolve the distinct species that contain more than three added G-nucleotides [9,23,44], but the pattern of migration observed here is consistent with the migration described previously for these species.

Interestingly, the addition of the external templating oligonucleotide did not appreciably affect the products observed in the reactions with ScThg1. Only a single  $G_{-1}$  residue was still observed, suggesting that ScThg1 is not using the multiple C-containing template that would result in multiple nucleotide addition, as previously observed for variant tRNA with this 3'-end sequence [9]. Instead, ScThg1 must be utilizing the native Sc-tRNA<sup>His</sup> 3'-end instead of the oligonucleotide-bound structure

as a template. This is possibly due to changes in the overall shape of tRNA<sup>His</sup> when bound to the oligonucleotide, which underscores this enzyme's stricter dependence on tRNA<sup>His</sup> structure that has been demonstrated repeatedly, and also the unsuitability of Thg1 members of the 3' to 5' polymerase family for this type of engineering approach.

In contrast, the addition of the external templating oligonucleotide to reactions with BtTLP results in significant changes in the observed reaction products. BtTLP efficiently created lower migrating products that have previously been identified as corresponding to multiple G-nucleotide additions [31], as expected for the use of the 3'-CCCCCA sequence as a template. This observation provides proof for the first time that TLPs are capable of accommodating an externally provided oligonucleotide, and also implies that the 3'-5' polymerase activity of TLPs can be guided to label an RNA substrate with a template of interest. Appearance of the products corresponding to multiple G-additions in the assays containing BtTLP supports the idea that BtTLP is an enzyme with more flexibility in accommodating an unnatural template and most importantly, BtTLP is capable of changing its nucleotide addition pattern to match the template that is available. The replacement of the tRNA<sup>His</sup> acceptor stem by the external oligonucleotide alone was sufficient to guide BtTLP to change its nucleotide addition preferences as opposed to Thg1.

## 5. Conclusions

The tRNA<sup>His</sup> guanylyltransferase family comprises a fascinating group of enzymes with novel catalytic activities. From the initial identification of these enzymes as the catalysts of adding a somewhat simple, albeit critical, single nucleotide to the 5'-ends of tRNA<sup>His</sup> species, the mechanistic and biological complexity that is associated with the members of this family has grown significantly. New surprises about the functions of these enzymes, both engineered and on natural RNA substrates, are sure to continue to emerge, and will provide new opportunities to take advantage of 3' to 5' polymerization in the future.

**Author Contributions:** A.W.C., M.I.J., I.U.H., and J.E.J. wrote the manuscript; M.I.J., C.Z.C., B.S.R., R.K., I.U.H., and J.E.J. prepared figures; B.S.R. carried out experiments; C.Z.C., M.I.J., I.U.H., and J.E.J. edited the manuscript.

**Funding:** This research was funded by Natural Sciences and Engineering Research Council of Canada grant number [RGPIN 04776-2014 to I.U.H.] and the J.P. Bickell Foundation (to I.U.H.), and the National Institutes of Health R01 GM087543 (to J.E.J.).

**Acknowledgments:** We are thankful to Patrick O'Donoghue for careful reading of the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Joyce, C.M.; Steitz, T.A. Polymerase structures and function: Variations on a theme? *J. Bacteriol.* **1995**, *177*, 6321–6329. [[CrossRef](#)] [[PubMed](#)]
2. Lehman, I.R.; Richardson, C.C. The deoxyribonucleases of *Escherichia coli*. Iv. An exonuclease activity present in purified preparations of deoxyribonucleic acid polymerase. *J. Biol. Chem.* **1964**, *239*, 233–241. [[PubMed](#)]
3. Uptain, S.M.; Kane, C.M.; Chamberlin, M.J. Basic mechanisms of transcript elongation and its regulation. *Annu. Rev. Biochem.* **1997**, *66*, 117–172. [[CrossRef](#)] [[PubMed](#)]
4. Hyde, S.J.; Eckenroth, B.E.; Smith, B.A.; Eberley, W.A.; Heintz, N.H.; Jackman, J.E.; Doublet, S. tRNA<sup>His</sup> guanylyltransferase (Thg1), a unique 3'-5' nucleotidyl transferase, shares unexpected structural homology with canonical 5'-3' DNA polymerases. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 20305–20310. [[CrossRef](#)] [[PubMed](#)]
5. Nakamura, A.; Nemoto, T.; Heinemann, I.U.; Yamashita, K.; Sonoda, T.; Komoda, K.; Tanaka, I.; Söll, D.; Yao, M. Structural basis of reverse nucleotide polymerization. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 20970–20975. [[CrossRef](#)]
6. Jahn, D.; Pande, S. Histidine tRNA guanylyltransferase from *Saccharomyces cerevisiae*. Ii. Catalytic mechanism. *J. Biol. Chem.* **1991**, *266*, 22832–22836.

7. Pande, S.; Jahn, D.; Söll, D. Histidine tRNA guanylyltransferase from *Saccharomyces cerevisiae*. I. Purification and physical properties. *J. Biol. Chem.* **1991**, *266*, 22826–22831. [[PubMed](#)]
8. Gu, W.; Hurto, R.L.; Hopper, A.K.; Grayhack, E.J.; Phizicky, E.M. Depletion of *Saccharomyces cerevisiae* tRNA<sup>His</sup> guanylyltransferase Thg1p leads to uncharged tRNA<sup>His</sup> with additional m(5)c. *Mol. Cell Biol.* **2005**, *25*, 8191–8201. [[CrossRef](#)] [[PubMed](#)]
9. Jackman, J.E.; Phizicky, E.M. tRNA<sup>His</sup> guanylyltransferase catalyzes a 3'-5' polymerization reaction that is distinct from G<sub>-1</sub> addition. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8640–8645. [[CrossRef](#)]
10. Jackman, J.E.; Phizicky, E.M. tRNA<sup>His</sup> guanylyltransferase adds G<sub>-1</sub> to the 5' end of trn<sup>his</sup> by recognition of the anticodon, one of several features unexpectedly shared with tRNA synthetases. *RNA* **2006**, *12*, 1007–1014. [[CrossRef](#)]
11. Ingle, R.A. Histidine biosynthesis. *Arabidopsis Book* **2011**, *9*, e0141. [[CrossRef](#)] [[PubMed](#)]
12. Lant, J.T.; Berg, M.D.; Heinemann, I.U.; Brandl, C.J.; O'Donoghue, P. Pathways to disease from natural variations in human cytoplasmic tRNAs. *J. Biol. Chem.* **2019**. [[CrossRef](#)]
13. Liao, S.M.; Du, Q.S.; Meng, J.Z.; Pang, Z.W.; Huang, R.B. The multiple roles of histidine in protein interactions. *Chem. Cent. J.* **2013**, *7*, 44. [[CrossRef](#)] [[PubMed](#)]
14. Sprinzl, M.; Vassilenko, K.S. Compilation of trna sequences and sequences of tRNA genes. *Nucleic Acids Res.* **2005**, *33*, D139–D140. [[CrossRef](#)]
15. Connolly, S.A.; Rosen, A.E.; Musier-Forsyth, K.; Francklyn, C.S. G-1:C73 recognition by an arginine cluster in the active site of *Escherichia coli* histidyl-tRNA synthetase. *Biochemistry* **2004**, *43*, 962–969. [[CrossRef](#)] [[PubMed](#)]
16. Nameki, N.; Asahara, H.; Shimizu, M.; Okada, N.; Himeno, H. Identity elements of *Saccharomyces cerevisiae* tRNA<sup>His</sup>. *Nucleic Acids Res.* **1995**, *23*, 389–394. [[CrossRef](#)]
17. Rosen, A.E.; Musier-Forsyth, K. Recognition of G<sub>-1</sub>:C73 atomic groups by *Escherichia coli* histidyl-tRNA synthetase. *J. Am. Chem. Soc.* **2004**, *126*, 64–65. [[CrossRef](#)]
18. Burkard, U.; Willis, I.; Söll, D. Processing of histidine transfer RNA precursors. Abnormal cleavage site for RNase P. *J. Biol. Chem.* **1988**, *263*, 2447–2451. [[PubMed](#)]
19. Orellana, O.; Cooley, L.; Söll, D. The additional guanylate at the 5' terminus of *Escherichia coli* tRNA<sup>His</sup> is the result of unusual processing by RNase P. *Mol. Cell Biol.* **1986**, *6*, 525–529. [[CrossRef](#)]
20. Cooley, L.; Appel, B.; Söll, D. Post-transcriptional nucleotide addition is responsible for the formation of the 5' terminus of histidine tRNA. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 6475–6479. [[CrossRef](#)]
21. Gu, W.; Jackman, J.E.; Lohan, A.J.; Gray, M.W.; Phizicky, E.M. tRNA<sup>His</sup> maturation: An essential yeast protein catalyzes addition of a guanine nucleotide to the 5' end of tRNA<sup>His</sup>. *Genes Dev.* **2003**, *17*, 2889–2901. [[CrossRef](#)] [[PubMed](#)]
22. Guo, D.; Hu, K.; Lei, Y.; Wang, Y.; Ma, T.; He, D. Identification and characterization of a novel cytoplasm protein Icf45 that is involved in cell cycle regulation. *J. Biol. Chem.* **2004**, *279*, 53498–53505. [[CrossRef](#)] [[PubMed](#)]
23. Rao, B.S.; Maris, E.L.; Jackman, J.E. tRNA 5'-end repair activities of tRNA<sup>His</sup> guanylyltransferase (Thg1)-like proteins from bacteria and archaea. *Nucleic Acids Res.* **2011**, *39*, 1833–1842. [[CrossRef](#)] [[PubMed](#)]
24. Yuan, J.; Gogakos, T.; Babina, A.M.; Söll, D.; Randau, L. Change of tRNA identity leads to a divergent orthogonal histidyl-tRNA synthetase/ tRNA<sup>His</sup> pair. *Nucleic Acids Res.* **2011**, *39*, 2286–2293. [[CrossRef](#)]
25. Wang, C.; Sobral, B.W.; Williams, K.P. Loss of a universal tRNA feature. *J. Bacteriol.* **2007**, *189*, 1954–1962. [[CrossRef](#)]
26. Rao, B.S.; Mohammad, F.; Gray, M.W.; Jackman, J.E. Absence of a universal element for tRNA<sup>His</sup> identity in *Acanthamoeba castellanii*. *Nucleic Acids Res.* **2013**, *41*, 1885–1894. [[CrossRef](#)] [[PubMed](#)]
27. Rao, B.S.; Jackman, J.E. Life without post-transcriptional addition of g-1: Two alternatives for trn<sup>his</sup> identity in eukarya. *RNA* **2015**, *21*, 243–253. [[CrossRef](#)] [[PubMed](#)]
28. Hyde, S.J.; Rao, B.S.; Eckenroth, B.E.; Jackman, J.E.; Doublet, S. Structural studies of a bacterial tRNA<sup>His</sup> guanylyltransferase (Thg1)-like protein, with nucleotide in the activation and nucleotidyl transfer sites. *PLoS ONE* **2013**, *8*, e67465. [[CrossRef](#)] [[PubMed](#)]
29. Kimura, S.; Suzuki, T.; Chen, M.; Kato, K.; Yu, J.; Nakamura, A.; Tanaka, I.; Yao, M. Template-dependent nucleotide addition in the reverse (3'-5') direction by Thg1-like protein. *Sci. Adv.* **2016**, *2*, e1501397. [[CrossRef](#)] [[PubMed](#)]

30. Lee, K.; Lee, E.H.; Son, J.; Hwang, K.Y. Crystal structure of tRNA<sup>His</sup> guanylyltransferase from *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.* **2017**, *490*, 400–405. [[CrossRef](#)] [[PubMed](#)]
31. Jackman, J.E.; Phizicky, E.M. Identification of critical residues for G<sub>-1</sub> addition and substrate recognition by tRNA<sup>His</sup> guanylyltransferase. *Biochemistry* **2008**, *47*, 4817–4825. [[CrossRef](#)] [[PubMed](#)]
32. Jackman, J.E.; Gott, J.M.; Gray, M.W. Doing it in reverse: 3'-to-5' polymerization by the Thg1 superfamily. *RNA* **2012**, *18*, 886–899. [[CrossRef](#)] [[PubMed](#)]
33. Smith, B.A.; Jackman, J.E. Kinetic analysis of 3'-5' nucleotide addition catalyzed by eukaryotic tRNA<sup>His</sup> guanylyltransferase. *Biochemistry* **2012**, *51*, 453–465. [[CrossRef](#)] [[PubMed](#)]
34. Betat, H.; Long, Y.; Jackman, J.E.; Morl, M. From end to end: tRNA editing at 5'- and 3'-terminal positions. *Int. J. Mol. Sci.* **2014**, *15*, 23975–23998. [[CrossRef](#)] [[PubMed](#)]
35. Desai, R.; Kim, K.; Buchsenschutz, H.C.; Chen, A.W.; Bi, Y.; Mann, M.R.; Turk, M.A.; Chung, C.Z.; Heinemann, I.U. Minimal requirements for reverse polymerization and tRNA repair by tRNA<sup>His</sup> guanylyltransferase. *RNA Biol.* **2018**, *15*, 614–622. [[CrossRef](#)]
36. Smith, B.A.; Jackman, J.E. *Saccharomyces cerevisiae* Thg1 uses 5'-pyrophosphate removal to control addition of nucleotides to tRNA<sup>His</sup>. *Biochemistry* **2014**, *53*, 1380–1391. [[CrossRef](#)] [[PubMed](#)]
37. Lant, J.T.; Berg, M.D.; Sze, D.H.W.; Hoffman, K.S.; Akinpelu, I.C.; Turk, M.A.; Heinemann, I.U.; Duennwald, M.L.; Brandl, C.J.; O'Donoghue, P. Visualizing tRNA-dependent mistranslation in human cells. *RNA Biol.* **2018**, *15*, 567–575. [[CrossRef](#)] [[PubMed](#)]
38. Preston, M.A.; D'Silva, S.; Kon, Y.; Phizicky, E.M. tRNA<sup>His</sup> 5-methylcytidine levels increase in response to several growth arrest conditions in *Saccharomyces cerevisiae*. *RNA* **2013**, *19*, 243–256. [[CrossRef](#)]
39. Rice, T.S.; Ding, M.; Pederson, D.S.; Heintz, N.H. The highly conserved tRNA<sup>His</sup> guanylyltransferase Thg1p interacts with the origin recognition complex and is required for the G2/M phase transition in the yeast *Saccharomyces cerevisiae*. *Eukaryot. Cell* **2005**, *4*, 832–835. [[CrossRef](#)]
40. Heinemann, I.U.; Randau, L.; Tomko, R.J., Jr.; Söll, D. 3'-5' tRNA<sup>His</sup> guanylyltransferase in bacteria. *FEBS Lett.* **2010**, *584*, 3567–3572. [[CrossRef](#)] [[PubMed](#)]
41. Heinemann, I.U.; Söll, D.; Randau, L. Transfer RNA processing in archaea: Unusual pathways and enzymes. *FEBS Lett.* **2010**, *584*, 303–309. [[CrossRef](#)]
42. Abad, M.G.; Long, Y.; Willcox, A.; Gott, J.M.; Gray, M.W.; Jackman, J.E. A role for tRNA<sup>His</sup> guanylyltransferase (Thg1)-like proteins from *Dictyostelium discoideum* in mitochondrial 5'-trna editing. *RNA* **2011**, *17*, 613–623. [[CrossRef](#)] [[PubMed](#)]
43. Heinemann, I.U.; Nakamura, A.; O'Donoghue, P.; Eiler, D.; Söll, D. tRNA<sup>His</sup>-guanylyltransferase establishes tRNA<sup>His</sup> identity. *Nucleic Acids Res.* **2012**, *40*, 333–344. [[CrossRef](#)]
44. Long, Y.; Abad, M.G.; Olson, E.D.; Carrillo, E.Y.; Jackman, J.E. Identification of distinct biological functions for four 3'-5' RNA polymerases. *Nucleic Acids Res.* **2016**, *44*, 8395–8406. [[CrossRef](#)]
45. Heinemann, I.U.; O'Donoghue, P.; Madinger, C.; Benner, J.; Randau, L.; Noren, C.J.; Söll, D. The appearance of pyrrolysine in trnahis guanylyltransferase by neutral evolution. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 21103–21108. [[CrossRef](#)] [[PubMed](#)]
46. Prat, L.; Heinemann, I.U.; Aerni, H.R.; Rinehart, J.; O'Donoghue, P.; Söll, D. Carbon source-dependent expansion of the genetic code in bacteria. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 21070–21075. [[CrossRef](#)]
47. Abad, M.G.; Rao, B.S.; Jackman, J.E. Template-dependent 3'-5' nucleotide addition is a shared feature of tRNA<sup>His</sup> guanylyltransferase enzymes from multiple domains of life. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 674–679. [[CrossRef](#)]
48. Dodbele, S.; Moreland, B.; Gardner, S.M.; Bundschuh, R.; Jackman, J.E. 5'-end sequencing in *Saccharomyces cerevisiae* offers new insights into 5'-ends of tRNA<sup>His</sup> and snoRNAs. *FEBS Lett.* **2019**, in press. [[CrossRef](#)]
49. Abad, M.G.; Long, Y.; Kinchen, R.D.; Schindel, E.T.; Gray, M.W.; Jackman, J.E. Mitochondrial tRNA 5'-editing in *Dictyostelium discoideum* and *Polysphondylium pallidum*. *J. Biol. Chem.* **2014**, *289*, 15155–15165. [[CrossRef](#)] [[PubMed](#)]
50. Dodbele, S.; Jackman, J.E.; Gray, M.W. Mechanisms and evolution of tRNA 5'-editing in mitochondria. In *RNA Metabolism in Mitochondria*; Jorge, C.-R., Gray, M.W., Eds.; Springer: Cham, Switzerland, 2018; Volume 34.
51. Gott, J.M.; Somerlot, B.H.; Gray, M.W. Two forms of RNA editing are required for tRNA maturation in *Physarum* mitochondria. *RNA* **2010**, *16*, 482–488. [[CrossRef](#)]

52. Lonergan, K.M.; Gray, M.W. Editing of transfer RNAs in *Acanthamoeba castellanii* mitochondria. *Science* **1993**, *259*, 812–816. [[CrossRef](#)] [[PubMed](#)]
53. Lonergan, K.M.; Gray, M.W. Predicted editing of additional transfer RNAs in *Acanthamoeba castellanii* mitochondria. *Nucleic Acids Res.* **1993**, *21*, 4402. [[CrossRef](#)] [[PubMed](#)]
54. Nakamura, A.; Wang, D.; Komatsu, Y. Molecular mechanism of substrate recognition and specificity of tRNA<sup>His</sup> guanylyltransferase during nucleotide addition in the 3'-5' direction. *RNA* **2018**, *24*, 1583–1593. [[CrossRef](#)] [[PubMed](#)]
55. Gao, Y.; Yang, W. Capture of a third Mg<sup>2+</sup> is essential for catalyzing DNA synthesis. *Science* **2016**, *352*, 1334–1337. [[CrossRef](#)]
56. Yang, W.; Weng, P.J.; Gao, Y. A new paradigm of DNA synthesis: Three-metal-ion catalysis. *Cell Bioscience* **2016**, *6*, 51. [[CrossRef](#)]
57. Yuan, Y.; Altman, S. Selection of guide sequences that direct efficient cleavage of mRNA by human ribonuclease P. *Science* **1994**, *263*, 1269–1273. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).