Spring 2020

Analysis of the role of RNA silencing protein 1 (Rsp1) in the biogenesis of ~23-24 nt sRNAs in Tetrahymena thermophila

Samantha Elizabeth O'Keffe Neff Ernest
Western Washington University, neffs4@wwu.edu

Follow this and additional works at: https://cedar.wwu.edu/wwuet

Part of the Biology Commons

Recommended Citation
Ernest, Samantha Elizabeth O'Keffe Neff, "Analysis of the role of RNA silencing protein 1 (Rsp1) in the biogenesis of ~23-24 nt sRNAs in Tetrahymena thermophila" (2020). WWU Graduate School Collection. 951.
https://cedar.wwu.edu/wwuet/951

This Masters Thesis is brought to you for free and open access by the WWU Graduate and Undergraduate Scholarship at Western CEDAR. It has been accepted for inclusion in WWU Graduate School Collection by an authorized administrator of Western CEDAR. For more information, please contact westerncedar@wwu.edu.
Analysis of the role of RNA silencing protein 1 (Rsp1) in the biogenesis of ~23-24 nt sRNAs in *Tetrahymena thermophila*

By

Samantha Elizabeth O’Keefe Neff Ernest

Accepted in Partial Completion of the Requirements for the Degree Master of Science

ADVISORY COMMITTEE

Dr. Suzanne R. Lee, Chair

Dr. Lynn Pillitteri

Dr. Anu Singh-Cundy

GRADUATE SCHOOL

David L. Patrick, Interim Dean
In presenting this thesis in partial fulfillment of the requirements for a master’s degree at Western Washington University, I grant to Western Washington University the non-exclusive royalty-free right to archive, reproduce, distribute, and display the thesis in any and all forms, including electronic format, via any digital library mechanisms maintained by WWU.

I represent and warrant this is my original work and does not infringe or violate any rights of others. I warrant that I have obtained written permissions from the owner of any third party copyrighted material included in these files.

I acknowledge that I retain ownership rights to the copyright of this work, including but not limited to the right to use all or part of this work in future works, such as articles or books.

Library users are granted permission for individual, research and non-commercial reproduction of this work for educational purposes only. Any further digital posting of this document requires specific permission from the author.

Any copying or publication of this thesis for commercial purposes, or for financial gain, is not allowed without my written permission.

Samantha Neff Ernest

April 30, 2020
Analysis of the role of RNA silencing protein 1 (Rsp1) in the biogenesis of ~23-24 nt sRNAs in *Tetrahymena thermophila*

A Thesis
Presented to
The Faculty of
Western Washington University

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science

by
Samantha Neff Ernest
April 2020
ABSTRACT

RNA interference (RNAi) pathways regulate a variety of biological processes, including normal cell growth and development, through the action of protein-RNA complexes containing small RNAs (sRNAs). Our research focused on an RNAi pathway in the ciliated unicellular eukaryote *Tetrahymena thermophila*. This pathway produces ~23-24 nucleotide (nt) sRNAs through the action of RNA-dependent RNA polymerase (RdRP) complexes (complexes termed RdRCs) and their interaction with an RNA nuclease called Dicer 2 (Dcr2). The accumulation of sRNAs also requires a protein called RNA Silencing Protein 1 (Rsp1) which associates with a subset of RdRC proteins. In this study, we first sought to learn more about the potential function and evolutionary conservation of Rsp1 by examining its sequence. Our results indicate Rsp1 may have structural similarity to RNA polymerases, including RdRPs, but lacks the conserved catalytic residues for RNA synthesis. We also identified Rsp1-like predicted proteins in other *Tetrahymena* species, but no clear homologs in more distantly related organisms.

Second, we tested three hypotheses for why Rsp1 is required for sRNA accumulation: 1) Rsp1 stabilizes the precursor RNA transcripts that are later processed into sRNA, 2) Rsp1 is necessary for the accumulation of RdRC proteins, and 3) Rsp1 is necessary for correct assembly of RdRCs. Our experimental results indicate that Rsp1 does not appear to regulate sRNA biogenesis by regulating the levels of sRNA precursors or RdRC proteins levels. Instead, purification of RdRCs revealed that in strains lacking Rsp1, RdRCs cannot be recovered. This suggests that RdRCs are disrupted somehow in the absence of Rsp1.
ACKNOWLEDGEMENTS

I would like to acknowledge my thesis advisor, Dr. Suzanne Lee, and my committee members, Dr. Lynn Pillitteri and Dr. Anu Singh-Cundy for their assistance and shared knowledge. I would also like to acknowledge the WWU Biology Department for providing support and funding for my research through the Thon Summer Fellowship, Biology Alumni Student Research Fellowship and Biology Faculty Fellowship as well as a teaching assistantship position. All these funding sources made my thesis research possible. I would also like to thank these undergraduate research assistants for their dedicated time and energy to this project, particularly, Chloe Jones, Brain Miller, and Madison Syverson. These students assisted with a variety of lab duties and without their help this masters research would not have been possible. I would also like to acknowledge contributions from WWU faculty: Dr. Dietmar Schwartz for assistance with phylogenetic analysis of Rsp1, Dr. Dan Pollard for assistance with statistical analysis of our Western blot data, Dr. John Antos for sharing his gel imaging equipment, Dr. Jeanine Amacher for training me with PyMol, and Dr. Filip Jagodzinski for aid with multiple sequence alignment and BLAST searches. In addition, I thank Dr. Kathy Collins (UC-Berkeley) for sharing unpublished research on the biochemical activities of Rsp1 to aid my research and for providing reagents for our use, and Dr. Dragony Fu (University of Rochester) for providing an alternative silver staining protocol. I would also like to thank the WWU Chemistry Department, especially Sam Danforth, for the use of their shared ultra-centrifuge. Many thanks to Sean Mckenna and the RiboWest conference committee for giving me the opportunity to discuss future directions for this research and financial support through a travel award and student poster award. Most of all, I would like to acknowledge my family and my husband, Kevin Ernest, for all of their morale, support, financial aid and kindness.
TABLE OF CONTENTS

Abstract ........................................................................................................................................... iv

Acknowledgements .......................................................................................................................... v

List of Tables and Figures ............................................................................................................... viii

Introduction ..................................................................................................................................... 1
  Background on RNA interference (RNAi) .................................................................................... 1
  Function and biogenesis of siRNA .............................................................................................. 2
  siRNA in *Tetrahymena thermophila* ......................................................................................... 4
  siRNA biogenesis in *Tetrahymena thermophila* ........................................................................ 5
  Insights into RNA-Silencing Protein 1 and experimental overview .......................................... 7
  My research questions and hypotheses ...................................................................................... 8

Methods ......................................................................................................................................... 10
  Bioinformatics analysis of Rsp1 ................................................................................................... 10
  *T. thermophila* culture handling ............................................................................................... 13
  *RSP1 KO* strain development .................................................................................................... 14
  RT-PCR experiments: strain confirmation and precursor transcript levels .............................. 16
Western blots and statistical analysis of tagged RdRC components .................................. 17

Affinity purifications to assess RdRC composition ......................................................... 18

Results ............................................................................................................................... 21

Analysis of Rsp1 amino acid sequence to identify potential conserved protein domains .......... 21

Identification of potential orthologs/homologs of Rsp1 in Tetrahymena species ............... 22

Identification of potential orthologs/homologs of Rsp1 beyond Tetrahymena ................. 24

Analysis of the predicted structural homology between Rsp1 and RdRPs ......................... 26

Assessment of RdRC-dependent sRNA precursor transcript levels with and without Rsp1 ...... 27

Analyzing a possible role of Rsp1 in RdRC component level regulation .......................... 28

RdRC composition analysis ............................................................................................. 29

Discussion .......................................................................................................................... 31

Rsp1 may be related to RdRPs but is unlikely to be catalytically active as an RdRP ...... 31

Rsp1 regulates the RdRC ................................................................................................. 35

Conclusions ...................................................................................................................... 37

Literature Cited .................................................................................................................. 39

Tables ................................................................................................................................ 44

Figures .............................................................................................................................. 48
LIST OF TABLES

Table 1. Rsp1 orthologs/homologs in *Tetrahymena* species. Blastp searches were performed in *Tetrahymena* comparative genome database using the Rsp1 amino acid sequence. Predicted proteins designated “Hypothetical Rsp1” were the top hits in all cases; *T. thermophila* Rdr1 and “Hypothetical RdRP” proteins were the second best hit by e-value. % identity reflects frequency of residue identity across BLAST aligned sequence regions. Reciprocal BLASTp results indicate whether or not a *Tetrahymena* BLASTp with the Rsp1 orthologs/homologs resulted in an Rsp1 or Rdr1 hit. Red box: Region in hit that aligned to Rsp1, Green box: RdRP domain as predicted by NCBI CDD 3.16 unless otherwise noted, red line: RdRP catalytic residues: DxDxD (not to scale), *: RdRP domain hit by SMART Pfam 31.0, *: RdRP domain hit by SMART Pfam 32.0........44

Table 2. Possible Rsp1 polymerase homologs in non-*Tetrahymena* species. Blastp searches were performed in a variety of databases. NCBI searches were restricted to the ciliates *Stentor, Oxytricha,* and *Paramecium,* respectively. Results from NCBI searches against *Tetrahymena* are included in the table for comparison. Color coding identical to Table 1 with the addition of a black bar indicating the DNA pol III predicted protein domain. Unc: unconfirmed protein, U.P.: UniprotKB, ParaDB: Paramecium Database, *O. trifallax* BLAST hit absent after 9/2017. .........................45

Table 3. Rsp1 predicted structural homology to RdRPs. Searches were performed using a variety of structural modeling algorithms and databases. Color coding and Table columns are identical to Table 1.................................................................46

Table 4. Quantification of RdRC component levels +/- Rsp1. Western blot bands were quantified using ImageStudioLite. Mean fold changes are for the normalized levels of the indicated protein in RSP1 knockout strains divided by the levels in wildtype backgrounds. Signals were normalized to either α-tubulin or total protein (Ponceau) stains, which served as sample loading controls. P-value was obtained using one-sample two-tailed t-tests. N: number biological replicates. ........................................................................................................................................47
LIST OF FIGURES

Figure 1: Depiction of a canonical siRNA biogenesis pathway. A single-stranded RNA transcript (red) is used as a template for double-stranded RNA synthesis by an RdRP (green) (in some eukaryotes) which may be found in complexes with other proteins (orange). Dicer (blue) recognizes the double stranded RNA and processes it into siRNA. The siRNA then binds an Argonaute family protein (grey) and functions to trigger silencing of gene expression by base-pairing to an RNA containing complementary sequence (purple).

Figure 2: Proposed model for of ~23-24 nt siRNA biogenesis in T. thermophila. Top: Shown to the left are the five types of predicted transcripts derived from the genomic loci that the most highly abundant siRNAs map to in clusters (Couvillion et al., 2009). These sRNAs are comprised of three Rdr1-dependent types (Pseudogene clusters I-V, High Copy Repeat, and Phased clusters) and one Rdr1-independent type (other endo-siRNAs). Different RNA-dependent RNA polymerase complexes (RdRCs, see A, left) are essential for the biogenesis of specific siRNAs that map to the different genomic clusters (Lee & Collins, 2006; Lee et al., 2009; Talsky & Collins, 2012). Bottom: A) the three known RdRCs, B) the two known RSPCs. Arrows indicate the apparent substitution of Rdr1 with Rsp1 in Rdn1-containing RSPCs (Lee, Talsky, & Collins, 2009; Talsky & Collins, 2012).

Figure 3. Tetrahymena Rsp1-like proteins contain a domain with amino acid sequence similarity but limited sequence identity with Tetrahymena RdRPs. MEGA alignments generated by MUSCLE of the Rsp1-aligning region from all Tetrahymena Rsp1-like and RdRP-like proteins listed in Table 1, with the exception of sequences from T. vorax and T. shanghaiensis, which had a high frequency of large indels which prevented their alignment. The region containing catalytic DxDxD residues for RdRPs is boxed in black, with consensus sequences/properties shown either on top or bottom of aligned sequences. Top) Alignment depicting sequence identity. Darker shading reflects higher frequency of residue identity across aligned sequences. Bottom) Alignment depicting conserved hydrophobicity (red)/hydrophilicity (blue) across aligned sequences.

Figure 4. Phylogenetic Analysis of putative Tetrahymena Rsp1-like and RdRP proteins. MEGA alignment of the full-length sequences of Tetrahymena Rsp1 orthologs and homologs, minus the T. shanghaiensis sequence due to large gaps in the alignment. Maximum Likelihood method and JTT matrix-based model with discrete gamma distribution (5 categories) used to model this tree, with branch lengths to scale. 0.5 scale bar indicated 50% genetic variation between branches of that scale size. Numeric code to the left of each sequence name refers to an NCBI or Tetrahymena Genome Database gene accession number.

Figure 5. Rsp1 has a domain with structural similarity to Neurospora crassa QDE1 but limited sequence homology. Phyre2 template-based model of Rsp1 against N. crassa QDE-1, 2j7n, aligned with N. crassa QDE-1 2j7o model (the identical structure). 100 atoms aligned in the “neck” region, RMSD of 0.256. A) QDE-1 shown in grey, Phyre2 predicted Rsp1 model in yellow. Red arrow depicts viewpoint shown in B). B) Close up of aligned RdRP “neck” region showing
structural similarity. C) Close up of aligned RdRP “neck” region showing identical residues in red, similar residues in orange, and non-conserved in grey. Corresponding sequence alignment shown below. Rsp1 amino acid range: 701-772 aa. ..........................................................................................................................50

Figure 6. Tetrahymena Rsp1-like proteins contain a domain with conserved amino acid sequence similarity but limited sequence identity with RdRPs beyond Tetrahymena species. Sequences selected are representative of the major branches from Figure 7. All Rdr1-like proteins within the same species as select Rsp1-like proteins were chosen to compare interspecies and intraspecies sequence similarity. Similar to Figure 3, the region containing catalytic DxDxD residues for RdRPs is boxed in black, with consensus sequences/properties shown either on top or on bottom of alignments, Rsp1 amino acids ranged from aa 1019-1122. Alignment regions with 90% sequence gaps were removed for ease of visualization. Top and bottom: MEGA alignment of Rsp1-like proteins and Rdr1-like with sequence identity or hydrophobicity/hydrophilicity, respectively, highlighted as in Figure 3. ..........................................................................................................................51

Figure 7. Phylogenetic Analysis of Tetrahymena Rsp1-like proteins and potentially related RdRP proteins. Tree was generated using the parameters described for Figure 4 using a MEGA alignment of T. thermophila Rsp1, putative Tetrahymena Rsp1 homologs, and RdRPs identified through BlastP and structural homology searches with T. thermophila Rsp1 and other RdRPs found in those same species. Sequences from A. alpina, C. sinensis, S. rodhaini, O. trifallax, T. vorax, and T. shanghaensis omitted due to large gaps in the alignment. Arrowheads mark sequences that were initially identified as BlastP or structural homology search hits with T. thermophila Rsp1. Red star indicates sequences that contain the DxDxD catalytic residues..........................................................52

Figure 8. RSP1 knockout strains confirmed with RT-PCR. RT-PCR performed on mRNA targets indicated on the right of gels, with actin mRNA as an internal positive control for RNA integrity. Δ: RSP1 knockout; MW: DNA ladder; gDNA: genomic DNA; RT: reverse-transcriptase; WT: parental strain containing tagged protein without ΔRSP1...........53

Figure 9. siRNA precursor transcript levels +/- Rsp1. RT-PCR performed on the siRNA precursor transcripts indicated on the right of gels, with actin mRNA as an internal positive control for RNA integrity. Δ: RSP1 knockout; MW: DNA ladder; gDNA: genomic DNA; RT: reverse-transcriptase; WT: parental strain containing tagged protein without ΔRSP1........................54

Figure 10. RdRC component levels +/- Rsp1. Western blots were performed using antibodies against the ZZ tag on A) Rdn1, B) Rdn2, and C) Rdr1 on extracts from growing cells (Rdn1 3xFZZ and Rdr1-ZZ) or cells starved for 6 hrs (Rdn2 3xFZZ) containing or lacking Rsp1 (ΔRSP1). Expression on Western blots was normalized to α-tubulin (A, B and C, top) or total protein as stained by Ponceau S (A and B, bottom). Two-fold dilutions of each sample were used identify signals in the linear range for quantification. MW: protein ladder; WT: parental strain with no tagged protein; *: cross reacting band in 6 hour starved cells. ...........................................................55

Figure 11. RdRC purifications +/- Rsp1. High stringency purifications were performed from overnight starved cultures. Bands corresponding in size to known RdRC components (and potential interacting partners) are denoted by the name of the component on the right side of the gel images. from tagged protein-expressing lines MW: protein ladder; WT: parental strain with no
tagged protein; TEV: TEV protease, used for RdRC elution from purification beads; ☆: known RdRC component; *: possible proteolytic Rdr1 product generated during sample collection.

Figure 12. Western blot analysis of Rdn1 in protein samples from Rdn1 3xFZZ affinity purification. A) Boiled whole cell samples collected prior to the preparation of extracts for affinity purification of Rdn1-containing RdRCs. B) “IP” samples collected after RdRC affinity purification on IgG beads. “Beads” samples represent IgG beads after RdRC elution from beads with TEV protease. *: cross reacting band.
INTRODUCTION

An overview of RNA interference (RNAi)

In diverse eukaryotes, from single-celled protists to vastly more complex multicellular eukaryotes, RNA interference (RNAi) pathways can regulate a wide variety of cellular processes. Among these RNAi-controlled processes are transcriptional and post-transcriptional gene regulation (Nishikura, 2001; Zamore & György, 2002), defense against viruses and transposable elements (Czech & Hannon, 2016; Obbard, Gordon, Buck, & Jiggins, 2009), and regulation of chromosome copy number (Khurana, Clay, Moreira, Wang, & Landweber, 2017). The study of RNAi pathways and associated biogenesis machinery has increased understanding of developmental biology (Garneau, Wilusz, & Wilusz, 2007; Tsuzuki, Motomura, Kumakura, & Takeda, 2017), led to advances in cancer research (Xu & Wang, 2015), and provided a conceptual starting point for the discovery of CRISPR-Cas systems from which new genome editing techniques have been developed. Although the biological function of RNAi pathways can vary between organisms, the fundamental component in all RNAi pathways is small non-coding RNA (sRNA), typically ranging from 17 to 30 nucleotides (nt) in size. The sRNAs that are best studied recognize their target DNA or RNA by complementary base pairing and in most cases, trigger heterochromatin formation or RNA degradation, respectively (Alberts et al., 2015; Flamand & Duchaine, 2012; Nishikura, 2001; Zamore & György, 2002). The three major classes of sRNAs that function in well-studied RNAi pathways are small interfering RNAs (siRNAs), micro-RNAs (miRNAs), and PIWI-associated RNAs (piRNAs). These three classes of sRNAs partially overlap in function, but are distinct in their biogenesis pathways (Carthew & Sontheimer, 2009; Czech & Hannon, 2016; Flamand & Duchaine, 2012). Despite all that has been learned about sRNAs, gaps
remain in our understanding, especially with respect to the diversity of sRNA types and functions, as well as how sRNA biogenesis is regulated. Here, I will focus on the biogenesis of siRNA-like sRNAs in the model organism *Tetrahymena thermophila*, a ciliated protist that has long served as a eukaryotic model for the study of many widely conserved cellular and molecular processes (Ruehle, Orias, & Pearson, 2016).

**Function and biogenesis of siRNA**

siRNAs can be distinguished from miRNAs and piRNAs by the long double-stranded RNA intermediates from which they derive (Carthew & Sontheimer, 2009; Czech & Hannon, 2016). These double-stranded intermediates are produced from precursor transcripts expressed from regions of the genome that are predicted to yield highly structured RNAs, act as mobile elements/transposons, or are otherwise not annotated as protein coding (Cerutti & Casas-Mollano, 2006; Chan, Yang, Huang, & Chang, 2013; Farley & Collins, 2017; Jung et al., 2016; Tschudi, Shi, Franklin, & Ullu, 2012; Vazquez et al., 2004). In certain eukaryotes, including *Tetrahymena thermophila* (the model eukaryote employed in my studies described here), this double-stranded RNA arises from a single-stranded RNA transcript which is then transformed into a double-stranded RNA transcript by the enzyme RNA-dependent RNA polymerase (RdRP) ([Figure 1](#), green) (Maida & Masutomi, 2011; Zamore & György, 2002; Zong, Yao, Yin, Zhang, & Ma, 2009). RdRPs sometimes exist in stable complexes, called RdRCs, with other proteins. Some of these RdRC components are nucleotidyltransferases, which function by adding one or more nucleotides to the 3’ end of a given transcript (also called TUTases, non-canonical poly(U) polymerases, or non-canonical poly(A) polymerases; [Figure 1](#), orange) (Martienssen, Zarrategui, & Goto, 2005). Once generated, the double-stranded precursor siRNA transcript is recognized and digested into
Figure 1: Depiction of a canonical siRNA biogenesis pathway. A single-stranded RNA transcript (red) is used as a template for double-stranded RNA synthesis by an RdRP (green) (in some eukaryotes) which may be found in complexes with other proteins (orange). Dicer (blue) recognizes the double stranded RNA and processes it into siRNA. The siRNA then binds an Argonaute family protein (grey) and functions to trigger silencing of gene expression by base-pairing to an RNA containing complementary sequence (purple).
smaller RNA fragments by the enzyme Dicer (Figure 1, blue) (Zamore & György, 2002). In some eukaryotes, RNA synthesis is directly coupled to double-stranded RNA processing, through a physical association of RdRP complexes with Dicer (Collins & Lee, 2007). Once generated, the siRNAs, like miRNAs and piRNAs, assemble with Argonaute family proteins to form the effector complex responsible for gene silencing or other RNAi-mediated activities (Carthew & Sontheimer, 2009). However, how RdRCs are assembled and identify their proper sRNA precursor substrates in the first place has remained unknown.

**siRNA in *Tetrahymena thermophila***

In *Tetrahymena thermophila*, two major classes of siRNAs have been described (Allen et al., 2017; Howard-Till & Yao, 2006; S. Lee & Collins, 2006). The first class of siRNAs to be identified in *T. thermophila* were ~28-30 nucleotide (nt) sRNAs termed “scan-RNAs” or scnRNA for their role in scanning parental nuclei and guiding the removal of DNA sequences that are not in the parental nuclei from the nuclei of progeny (Mochizuki & Gorovsky, 2004). This now highly-studied pathway was one of the first identified mechanisms for RNA-mediated transgenerational inheritance in any organism (Allen et al., 2017; Chalker, Meyer, & Mochizuki, 2013; Coyne, Lhuillier-Akakpo, & Duharcourt, 2012). The second major class of siRNAs in *T. thermophila* – and the focus of my research described here - is approximately 23-24 nt in size. The biological processes that the ~23-24 nt sRNAs regulate remain unknown, though some studies have suggested that ~23-24 nt sRNAs may function in either transcriptional or post-transcriptional gene silencing (Farley & Collins, 2017; Howard-Till & Yao, 2006; Lee & Collins, 2006).

The most abundant *T. thermophila* ~23-24 nt sRNAs can be further subdivided into four different subclasses, based on their genomic origins as determined by high-throughput sequencing. These subclasses are distinguished by the types of discrete genomic loci that the sRNA sequences
map to: pseudogene-/retroelement-like clusters, high copy repeat regions, regions that are predicted to be transcribed to yield RNAs with distinctive stem-loop structures (phased hairpin clusters or Twi8-associated stem loops, known as TASLs), and regions predicted to give rise to double-stranded RNA or encode protein (endo-siRNA) (Figure 2, top; Couvillion et al., 2009; Farley & Collins, 2017; Talsky & Collins, 2012). SiRNA-producing clusters from similar types of genomic origins can be found in a variety of other eukaryotes, suggesting that comparable siRNA biogenesis pathways may exist in other species (Cerutti & Casas-Mollano, 2006; Chan, Yang, et al., 2013; Chan, Yuo, et al., 2013; Farley & Collins, 2017; Jung et al., 2016; Tschudi et al., 2012; Vazquez et al., 2004).

siRNA biogenesis in *Tetrahymena thermophila*

Insights into the biogenesis pathway of the *T. thermophila* ~23-24 nt siRNAs first came through biochemical analysis of the only known RdRP in *Tetrahymena*, Rdr1 (Collins & Lee, 2007). Through *in vitro* assays, affinity purified complexes containing Rdr1 were found to generate both double-stranded RNA and ~23-24 nt siRNAs from single-stranded RNA transcripts. Proteomic mass spectrometry revealed that Rdr1 exists in a salt- and heat-stable RdRC containing one or two additional proteins (Figure 2, bottom, part A) and associates in a salt- and heat-sensitive manner with Dcr2, one of three Dicer homologs in *Tetrahymena* (Figure 2, top) Collins & Lee, 2007). Notably, the *in vitro* production of ~23-24 nt siRNAs by purified RdRCs was blocked by disruption of the interaction between RdRC and Dcr2. In addition to Rdr1, the RdRCs...
Figure 2: Proposed model for of ~23-24 nt siRNA biogenesis in *T. thermophila*. Top: Shown to the left are the five types of predicted transcripts derived from the genomic loci that the most highly abundant siRNAs map to in clusters (Couvillion et al., 2009). These sRNAs are comprised of three Rdr1-dependent types (Pseudogene clusters I-V, High Copy Repeat, and Phased clusters) and one Rdr1-independent type (other endo-siRNAs). Different RNA-dependent RNA polymerase complexes (RdRCs, see A, left) are essential for the biogenesis of specific siRNAs that map to the different genomic clusters (Lee & Collins, 2006; Lee et al., 2009; Talsky & Collins, 2012). Bottom: A) the three known RdRCs, B) the two known RSPCs. Arrows indicate the apparent substitution of Rdr1 with Rsp1 in Rdn1-containing RSPCs (Lee, Talsky, & Collins, 2009; Talsky & Collins, 2012).
contain one of two RNA-dependent nucleotidyltransferases (Rdn1 or Rdn2), which add uridine-rich tails to the 3’ end of single stranded RNAs in vitro, though the biological function of this U-tailing activity in ~23-24 nt siRNA biogenesis remains unclear (Collins & Lee, 2007; Lee, Talsky, & Collins, 2009; Talsky & Collins, 2012). RdRCs containing Rdn1 also contain one of two additional proteins, Rdf1 or Rdf2, that lack any known domains.

Evidence that ~23-24 nt siRNAs found in vivo are processed by Rdr1 came from the finding that siRNAs derived from pseudogene-/retroelement-like clusters, high copy repeats and phased hairpin/TASL clusters fail to accumulate in strains induced to overexpress a catalytically-inactive version of Rdr1 (Talsky & Collins, 2012). Moreover, different populations of Rdr1-dependent ~23-24 nt sRNAs fail to accumulate in T. thermophila strains bearing knockouts of RDN2, RDF1, and RDF2, suggesting that each RdRC is responsible for the biogenesis of distinct subclasses of these siRNAs, as outlined in Figure 2, top (Couvillion et al., 2009; Farley & Collins, 2017; Talsky & Collins, 2012). The biogenesis of the endo-siRNA appears to be Rdr1-independent: accumulation of this type of siRNA was not reduced in the absence of RdRC components or through overexpression of catalytically-inactive Rdr1(Couvillion et al., 2009; Talsky & Collins, 2012). However, given the ~23-24 nt size of this class of siRNAs, endo-siRNA biogenesis may still be dependent on Dcr2 (Figure 2, top).

**Insights into a novel sRNA biogenesis factor, RNA-Silencing Protein 1**

An additional protein, RNA silencing protein 1 (Rsp1), was identified in affinity purifications of Rdn1 and appears to replace Rdr1 in Rdn1-containing complexes termed RNA silencing protein complexes (RSPCs; Figure 2, bottom, part B) (Lee et al., 2009; Talsky & Collins, 2012). For reasons not yet identified, Rdn1 activity as a U-tailing enzyme is suppressed in RSPC
complexes (Talsky & Collins, 2012). In addition, unlike RdRCs, RSPCs do not appear to interact with Dcr2. These findings suggest that RSPCs are not functionally redundant to RdRCs. Yet, knockouts of Rsp1 fail to accumulate all Rdr1-dependent siRNAs, suggesting that Rsp1 is somehow critical for the biogenesis of these siRNAs (Talsky & Collins, 2012). Moreover, Rdr1-independent endo-siRNAs appear to be misregulated in RSP1 gene knockout strains, accumulating with siRNA lengths longer than the typical ~23-24 nt. Notably, this curious alteration in endo-siRNA length was also observed in strains that overexpress the catalytically inactive mutant of Rdr1 (Talsky & Collins, 2012). Together, these data indicate that Rsp1 somehow contributes to the normal functions of RdRCs and perhaps Dcr2 in siRNA biogenesis, yet the mechanism for this remains unknown.

My research questions and hypotheses

Given the lack of a molecular understanding of how Rsp1 contributes to the ~23-24 nt sRNA RNAi pathway in T. thermophila, my research thus focused on exploring the possible functions of Rsp1 in the biogenesis of sRNAs in T. thermophila. Specifically, the questions guiding my study were the following:

Question I. Does Rsp1 have any conserved domains? If so, do these predicted domains suggest possible functions Rsp1 may serve?

Question II. Does Rsp1 have any homologs? If so, are any of these homologs characterized, thereby suggesting possible functions Rsp1 may serve?

Question III. Is Rsp1 necessary for precursor transcript accumulation?

Question IV. Is Rsp1 necessary for RdRC component accumulation?

Question V. Is Rsp1 necessary for proper RdRC composition?
In pursuing answers to each of the questions listed above, we have gained a new understanding of the evolutionary relationship between Rsp1 and RdRPs and refined our model for the possible roles of Rsp1 in the *T. thermophila* sRNA biogenesis pathway.
METHODS

Bioinformatics analysis of Rsp1

Identification of potential Rsp1 homologs and orthologs. The amino acid sequence of full-length Rsp1 and the RdRP-like domain identified by SMART analysis was analyzed by BLASTp and tBLASTn algorithms on the following web databases: NCBI (BLAST+2.8.0-alpha)(Marchler-Bauer et al., 2017), Ciliate.org (referencing all ciliate genomes available as of October 2018), *Tetrahymena* comparative genome database (27 February 2019 version)(Yang et al., 2019), ParameciumDB (2011 version, old and new)(Arnaiz, Cain, Cohen, & Sperling, 2007; Arnaiz & Sperling, 2011), Dictyostelium DB (6 August 2012 version)(Fey, Dodson, Basu, & Chisholm, 2013), GiardiaDB (version 43)(Aurrecoechea et al., 2009), TAIR (version 2.2.80(Berardini et al., 2015), WormBase (version WS268)(Lee et al., 2018), Saccharomyces Genome Database (version S288C)(Chervitz et al., 1999), NeurosporaDB (version 43, OR74A v2.0)(Nordberg et al., 2014), EupathDB (version 43)(Aurrecoechea et al., 2017), reefDB (version 2.7.6), ToxoplasmaDB (version 43)(Gajria et al., 2008), EnsemblFungi with Pombase (version 45)(Lock et al., 2019), and TrypanosomaDB (version 43)(Aslett et al., 2009). All databases were last accessed June 2019 with the expect value set to 50. NCBI searches were completed using both the default parameters and the following altered parameters: the expect value was increased to 50 (rather than the default 10) and the word size was decreased to 3 (rather than the default 6). These altered parameters for the NCBI searches were determined by trial and error to be the most constrained parameters that would still identify both Rsp1 and Rdr1 in searches. NCBI searches were completed using all databases available including or excluding the following model ciliates and other unicellular eukaryotes with relatively well annotated genomes: *Ichthyophthirius multifilis, Stentor coeruleus, Oxytricha*
trifallax, Paramecium tetraurelia, Stylonychia lemnæ, Plasmodium falciparum, Giardia lamblia, Toxoplasma gondii, Schizosaccharomyces pombe, and Trypanosoma brucei.

Reciprocal analysis of each of the top hits (ie. results with the lowest e-values) from the BLASTp and tBLASTn searches was then executed through NCBI to examine the strength of homolog relatedness to Rsp1. For these searches, the homologous sequence was analyzed by NCBI BLASTp against Tetrahymena database entries with the expect value increased to 50 and word size adjusted to 3. Reciprocal analysis was also completed through Ciliate.org with the expect value increased to 50 using putative Rsp1 ortholog protein sequences identified in eight out of ten sequenced Tetrahymena species. Searches of the sequenced genomes of two Tetrahymena species, T. paravorax and T. shanghaiensis, did not reveal a clear candidate for an Rsp1 ortholog and were therefore not included in this reciprocal analysis.

**Rsp1 homolog amino acid alignments.** Sequence alignments were generated in MEGAX by MUSCLE with a Multiple Alignment Gap Opening penalty set to 2.9, Hydrophobicity Multiplier set to 1.2, Max Memory in MB set to 2048, Max iterations set to 16, Cluster Methods set to UPGMA, and the Multiple Alignment Gap Extension penalty set to 1.8. Amino acid sequence regions that failed to align for greater than 90% of the homologs were removed using the MEGA alignment editor tools (Hall, 2013). These alignments were then opened in UGENE and annotated with different colors to represent sequence identity and hydrophobicity trends.

**Phylogenetic tree building.** Phylogenetic trees were built using MEGAX MUSCLE sequence alignments with the following settings: Statistical Method – Maximum Likelihood, Test of Phylogeny – Bootstrap method, Number of Bootstrap Replications – 500, Substitution Model – Jones-Taylor-Thornton (JTT) model, Rates among Sites – Gamma Distributed (G), Number of
Assessment of Rsp1 Domains by amino acid sequence (primary structure). The full-length amino acid sequence of Rsp1, *RSP1_Tthermophila*, was analyzed using the following online domain analysis algorithms: National Center for Biotechnology Information conserved domain database (NCBI CDD; version 3.16), and Simple Modular Architecture Research Tool (SMART; version 8.0)(Letunic, Doerks, & Bork, 2015). NCBI CDD and SMART analysis was completed using all default search parameters. SMART searches referenced the following databases: P-FAM domains (version 31.0 and 32.0), HMMER, SignalP 4.1 server, and outlier homolog profiles provided by Steffen Schmidt (Schmidt, Bork, & Dandekar, 2002).

Assessment of predicted higher order structure homology for Rsp1. Using all default settings unless otherwise noted, the full length amino acid sequence of Rsp1 was analyzed using the following online structural analysis algorithms: Protein Homology/analogY Recognition Engine (Phyre2; version 2.0)(Kelley, Mezulis, Yates, Wass, & Sternberg, 2015), Bioinformatics Toolkit HHblits and HHpred (version v32.0)(Zimmermann et al., 2018), Protein Model Portal (PMP; version 6808)(Haas et al., 2013), Comparison of Multiple Protein sequence Alignments with assessment of Statistical Significance (COMPASS; version 9 September 2008)(Sadreyev, Tang, Kim, & Grishin, 2009), PredictProtein (28 January 2012 version)(Yachdav et al., 2014), Protein Structure Prediction Server (PS2; version 3.0)(Chen, Hwang, & Yang, 2009), and RaptorX (2013 version)(Källberg et al., 2012). All algorithms were last accessed June 2019. Phyre2 searches were completed using both normal and intensive modes. HHblits searches were completed using both
uniprot20_2016_02 and unicluster30_2018_08 databases. HHpred searches were completed using all available databases as of October 2018, which included SCOPe70_2.07, NCBI_Conserved_Domain(CD)_v3.16, SMART_v6.0, TIGR_FAMs_v15.0, PRK_v6.9, COG_KOG_v1.0, Pfam-A_v31.0, PRK_v6.9, PDB_mmCIF70_5_Oct, PDB_mmCIF30_5_Oct, ECOD_F70_20180219.

Protein sequences identified from these analyses (termed “Rsp1 homologs”) were then analyzed by NCBI CDD and SMART as described above in “Assessment of Rsp1 Domains.” Rsp1 homologs that had a predicted RdRP domain were aligned with known RdRPs such as rrf-3 (C. elegans), RdRP1 (A. thaliana), and Rdr1 (T. thermophila) using the multiple sequence alignment software Clustal Omega (version 1.2.2) in order to locate any conserved RdRP catalytic residues (Sievers et al., 2011).

**T. thermophila culture handling**

*T. thermophila* cells were grown in 1 mL room temperature working stock cultures in 24-well plates and passaged by pipetting 30-500 μL of dense cultures into fresh *Tetrahymena* growth media, SUPP (8 g/L proteose peptone, 0.8 g/L bacto-yeast extract, 12 μM iron, 0.2% glucose, 250 μg/mL Ampicillin, 250 μg/mL Streptomycin, and a 1:200 dilution of Fungizone) every 4-7 days, as needed. To prepare larger vegetatively growing cultures for analyses, working stock cultures were used to make 3 mL cultures in SUPP grown at 30°C, which could then be used to establish even larger cultures or used directly for analyses. The volume of cells added varied based on the strain, density and health of the culture stocks. Cultures were grown overnight in a shaking incubator at 125-150 rpm, 30°C to early-mid-log densities (~2.5-4 x 10⁵ cells/mL). Starved cultures were prepared by gently pelleting cells from vegetative cultures at 225xg for 3 minutes at room
temperature, washing twice in 10mM Tris pH 7.5, and resuspending in Tris for continued shaking at 125-150 rpm, 30°C.

**RSP1 KO strain development**

**Optimization of drug selection conditions for biolistic transformation.** An existing *RSP1* knockout (KO) construct which carries the *BSR2* gene cassette designed to confer resistance to blasticidin in the presence of Cd$^{2+}$ was used to generate new *RSP1* KO strains. To initially establish a concentration of blasticidin that could be used to select for desired transformation with the *RSP1* KO construct, cells of the desired genetic background were grown to mid-log ($\sim$2.5-4 x $10^5$ cells/ml) and starved for 24 hours, while shaking at 30°C. Cells were then adjusted to $\sim$2.0 x $10^5$ cells/mL and placed in the 30°C incubator without shaking to recover for 4 hours to mimic selection conditions. Cells were then selected in wells of a 96-well plate with a range of blasticidin concentrations in growth media that spanned final concentrations of 0 to 800 $\mu$g/mL and Cd$^{2+}$ at a final concentration of 0.5 $\mu$g/mL. Cells were incubated at 30°C in a humidified chamber and checked daily for cell death. The lowest blasticidin concentration that successfully killed non-transformed cells was identified as the appropriate starting concentration to use for transformant selection. For Rdn1 3xFZZ and Rdn2 3xFZZ, 400 $\mu$g/mL blasticidin and 0.5 $\mu$g/mL Cd$^{2+}$ and incubation for 4 days was identified as sufficient for initial selection conditions.

**Transformation of Tetrahymena with the RSP1 KO construct.** Transformations were performed according to standard *Tetrahymena* protocols using a biolistic transformation apparatus (Biolistic PDS-1000/He Particle Delivery System, Bio-Rad, Cassidy-Hanley, 2012). Briefly, 20 $\mu$L of 50 $\mu$g/mL Seashell gold (S550d) was transferred to an Eppendorf tube and mixed with 13
µL binding buffer and 2 µg linearized DNA by vortexing for incubation at room temperature for 3 minutes. Next, the DNA coated gold beads were then collected by microcentrifugation at 9400 xg for 10 seconds, rinsed once with 500 µL of 100% ice-cold EtOH, and resuspended in 20 µL EtOH. Resuspended beads were finally evenly spread onto the macrocarriers in holders and desiccated for 10 minutes.

In preparation for transformation, 50 mL of 24-hour starved cells were collected at 1100xg for 2 minutes and the resuspended cell pellet was gently pipetted onto Tris-moisturised filter paper in a petri plate lid. These cells were slid into the second shelf of transformation apparatus and bombarded at ~900-1000 psi with the DNA-coated gold beads under vacuum. Cells were recovered with 50 mL pre-warmed SUPP lacking Cd2+ and blasticidin and transferred to a flask at 30°C (not shaking) for 4 hours for recovery.

**Gene knockout selection.** For selection of transformants, bombarded cells were plated into a 96 well plate with each well containing a 2:1 ratio of recovered cells to fresh media containing SUPP/Cd²+/blasticidin, with starting concentrations of Cd²+ and blasticidin based on the blasticidin optimization described above. These plates were transferred to a dark humidified chamber box at 30°C without shaking for four days. Any wells with surviving cells were then passaged roughly every other day for several weeks, with deviations in passage frequency of a single day dependent on apparent culture density. Blasticidin concentrations were increased while Cd²+ concentrations were decreased with every passage until maximum selective conditions that still permitted culture growth were reached. Once cells were passaged at the maximum selective conditions for 3 weeks, individual cells were pipetted into fresh wells, passaged in SUPP lacking blasticidin and Cd²+ for two weeks, and grown up in larger cultures for genotype verification by RT-PCR.
RT-PCR experiments: strain confirmation and precursor transcript levels

**Total RNA preparation and DNase treatment.** At least 14 candidate *RSP1 KO* strains per desired genotype were grown up at 30°C in the shaking incubator until ~2-5 x 10^5 cells/mL. These cultures were then harvested at 225xg at room temperature in a tabletop centrifuge, and cell pellets were immediately resuspended in 500 μL Trizol (Invitrogen) and transferred to the -20°C freezer for storage. RNA preparation was resumed according to the manufacturer’s protocol. RNA pellets were initially resuspended in 35μL RNase-free water for concentration measurements using a Nanodrop1000, and concentrations were adjusted, if needed, to ~1 μg/μL RNA. RNA was then treated with RNase-free DNase I (RNase free; NEB) for 1hr at 37°C. To remove DNase I, RNA was extracted with 25:24:1 phenol-chloroform-iso-amyl-alcohol (PCI), precipitated using 3M sodium acetate (pH 5.2), and rinsed with 70% RNase-free ethanol. The dried RNA pellet was resuspended to 1 μg/μL in RNase-free water.

**Reverse transcription and polymerase chain reaction.** For reverse transcription (RT) reactions, 10 μL reactions volumes were used that contained 1.75 μg/μL RNA, 3.5 x 10^-6 mmol primer, 3.5 x 10^-5 mmol dNTPs, 20% v/v Superscript IV 5x buffer, 5 x 10^-8 mol DTT, 50 units Superscript and 15 units RNase out. “-RT” negative control reactions were included that lacked Superscript IV. Reactions were carried out for 90 minutes at 50°C followed by 10 minutes of heat-inactivation at 80°C. Reactions were then treated with 5 units RNase H and incubated 20 minutes at 37°C to remove RNA hybridized to cDNA.

One μL of RT reaction products was then used for 25 μL PCR reactions which each contained final concentrations of: 1X Standard Taq Reaction Buffer, 5.0 x 10^-6 mmol dNTPs, 5.0
x $10^6$ umol per primer, 0.625 units Taq DNA Polymerase. A water control lacking RT products as a template was included as a negative control reaction.

**Western blots and statistical analysis of tagged RdRC components**

**Culture growth and sample harvesting.** For Westerns performed on whole cell lysates, $\sim 1 \times 10^6$ cells were harvested from mid-log or starved cultures for each Western blot sample. Cells were rinsed once in 10mM Tris and harvested or harvested directly without a Tris rinse at 200-225xg or 500xg for 3 minutes, for vegetative or starved cultures respectively. Once pelleted, ice-cold lysis buffer was added at 6 times the pellet volume to the cells (10% Glycerol, 20mM Tris, pH 8.0, 50mM NaCl, 1mM MgCl$_2$, 1x Protease inhibitor cocktail (Sigma), 0.1mM $\beta$-mercaptoethanol ($\beta$-Me), and 0.1 mM PMSF. Resuspended cells were rocked end-over-end at 4°C for 15 minutes, and then boiled at 100°C for 5 minutes. After boiling, the samples were cooled to room temperature and 5x SDS-PAGE sample buffer (0.3M Tris-HCl pH 6.8, 10% SDS, 20% $\beta$-Me, 50% glycerol, bromophenol blue) was added to a final concentration of 1x. All samples were stored at $-20^\circ$C.

In preparation for loading on SDS-PAGE gels, samples were fully thawed, vortexed, boiled for 3 minutes and pelleted at 17,000xg for 5 minutes. Samples collected during affinity purifications described below were prepared similarly, without boiling.

**Western blotting.** Samples were run on 9% SDS-PAGE gels and transferred onto nitrocellulose for 0.2% Ponceau staining, followed by blocking in 5% powdered non-fat milk in TBS (6.35 g/L Tris HCl, 1.18 g/L Tris base, 8.76 g/L NaCl) for 45-120 minutes. Primary antibody incubations were carried out at 4°C overnight, rocking gently, with 1:500 diluted rabbit IgG (Sigma, I5006) to detected ZZ-tagged proteins (Rdn1, Rdn2 and Rdr1) or 1:620 diluted mouse anti-α-tubulin (Sigma,
T6793-100 µL) in 5% milk/TBS. Blots were washed in TBS once for 15 minutes, followed by three 5-minute rinses. Secondary antibodies in 5% milk/TBS (1:15,000 diluted Donkey anti-Rabbit or Goat anti-Mouse) were incubated with blots for 1-3 hours at room temperature, followed by the same blot TBS washes described above. Blots were developed in ECL Prime (manufacture and catalog number) before imaging using the LI-COR Odyssey Fc with a 10 minute “Chemi” channel (to visualize the secondary antibodies) and 30 second “700” channel (to visualize the protein ladder bands).

**Quantification and statistical analysis of Western blots.** ImageStudioLite was used to quantify Western blot signals. Initially, all signals were assessed to verify that they were within a linear range as determined from a dilution series of protein sample samples and two-fold over background. Background signals were then subtracted from signals that satisfied these criteria. Next, the fold-change of IgG signals normalized to α-tubulin or Ponceau signals in KO strain samples compared to WT strain samples was determined. Fold-change values were logarithmically transformed for 3-4 replicates for statistical analysis using a one-sided two-tailed T-test.

**Affinity purifications to assess RdRC composition**

**Preparation of S100 cell extracts.** Large-scale cultures (500ml to 1L) grown to ~3.5-5 x 10^5 cells/mL were used for purifications. Briefly, cells were harvested by centrifugation at 1000xg for 2 minutes, and washed once in 10mM Tris, pH 7.5 before resuspending them in three-fourths of the culture volume and returned to 30°C for shaking overnight. The same number of cells from each strain was then recovered by centrifugation at room temperature at 1500xg.
Harvested cells were lysed with 3 times the pellet volume of lysis buffer (20mM Tris-HCl pH 8.0, 1mM MgCl₂, 10% glycerol, 50mM NaCl, with freshly added 1mM β-Me, 0.1 mM PMSF, 1x Protease inhibitor cocktail (Sigma P8340 used as a 1000x stock), and 1x Aprotinin). Immediately after adding the lysis buffer and resuspending the cell pellet, the detergent IGEPAL was added to a final concentration of 0.2% and lysates were rotated end-over-end for 15 minutes at 4°C. Lysates were clarified by spinning at 4°C 100,000xg for 1 hour in ultraclear ultracentrifuge tubes (SW32.1 rotor, Beckman conical tubes #358123). After the spin, the supernatant (S100) was transferred to a clean conical tube and additional PMSF was added to a final concentration of 0.1 mM.

**Affinity purification, bead washes, and TEV elution.** Rabbit IgG agarose beads were prepared by rinsing with room temperature T2MGN50 buffer (20mM Tris-HCl pH 8.0, 1mM MgCl₂, 10% glycerol, and 50mM NaCl) three times, recovering beads between rinses at 1500xg. Affinity purifications were performed with 2 μL packed beads per mL of S100 and transferred to 4°C to rotate end-over-end for 1.5 hours. Beads were then rinsed with ice-cold T2MGN50 containing 0.1 mM PMSF, divided into ~15 μL packed bead volumes among low-retention Eppendorf tubes, and washed three times for 5 minutes with room temperature T2MGN200 (200mM NaCl rather than the 50mM NaCl used above) containing 0.1 mM PMSF. After three washes, beads were transferred to fresh low-retention tubes and washed three times with T2MGN200 lacking PMSF. To elute purified complexes, beads were incubated 1 hour at 4°C rotating end-over-end in 4x bead volumes of T2MGN50 buffer containing homemade TEV protease (gift of K. Collins, UC-Berkeley). A second elution was often performed overnight to recover as much material as possible. All TEV eluates and beads post-elution were stored in 1x SDS-PAGE Sample buffer lacking β-Me at -20°C.
**Silver staining.** To visualize proteins in affinity purification samples, samples were run on 9% SDS-PAGE gels without prior boiling or addition of β-Me. SDS-PAGE gels were stained for total protein in purified samples as follows in acetone-cleaned glass bowls at room temperature. Gels were fixed in 40% ethanol, 10% acetic acid for 1 hour to overnight, rocking gently. Following fixation, gel was incubated for 30 minutes in sensitizing solution (30% ethanol, 0.2% sodium thiosulphate, 6.8% sodium acetate in ddH2O) and then rinsed in deionized and distilled water. To stain, gels were incubated for 20 minutes in the dark with 0.25% silver nitrate prepared in nanopure water. Following a rinse in nanopure water, stain was developed in 2.5% sodium carbonate, 0.015% formaldehyde until protein bands were clearly visible by eye over a lightbox. Developing was stopped with a one-minute incubation of 5% acetic acid solution followed by rinses in water. Gels were imaged using the BIO-RAD Gel Doc EZ Imager with Image Lab 5.2.1 and incubated briefly in 10 mM Tris pH 7.5 before drying on a benchtop at room temperature between sheets of cellophane.
RESULTS

Analysis of Rsp1 amino acid sequence to identify potential conserved protein domains

A previous study of Rsp1 had not identified any conserved protein domains in the protein sequence (Talsky & Collins, 2012); however, given the improvement of search tools and databases since the initial study, we sought to revisit this analysis for clues to Rsp1 molecular functions in sRNA biogenesis. To do this, I used two publicly available domain prediction programs: Simple Modular Architecture Research Tool (SMART) and the Conserved Domain Database at the National Center for Biotechnology Information (NCBI CDD). SMART functions by first performing multiple sequence alignments between a query sequence with representative modeled domain family members, based on structurally analyzed proteins and predicted models of proteins. These alignments are optimized by a series of algorithms and a hidden Markov model is developed for further searches with HMMer, Pfam or other databases (Schultz, 2000). NCBI CDD identifies predicted protein domains by generating alignment models between a query sequence and representative protein domain family members, which agree with experimentally confirmed tertiary structures (Marchler-Bauer et al., 2017). Using the previously described full-length amino acid sequence of Rsp1 as a query (Talsky & Collins, 2012), which was deposited in the Tetrahymena Genome Database (Yang et al., 2019), I found that Rsp1 bears an amino acid region with intriguing homology to a cellular RNA-dependent RNA polymerase (RdRP) domain.

Cellular RdRPs function by processing single stranded RNA into double stranded RNA and are structurally distinct from viral RdRPs despite their similar biochemical activities (Maida & Masutomi, 2011; Qian et al., 2016; Venkataraman, Prasad, & Selvarajan, 2018). The SMART Pfam Rsp1 RdRP domain result extends between residues 686-861 with an e-value of 2.4 x 10^-6.
A similar result was obtained using NCBI CDD, with a predicted RdRP domain in Rsp1 extending between residues 738-816 with an e-value of 2.0 x 10^{-2}. Importantly, however, functional cellular RdRP domains characteristically have a “DxDxD” sequence, which is believed to coordinate a Mg^{2+} ion enabling the addition of a new nucleotide to the nascent strand of an RNA molecule (Salgado et al., 2006). These “catalytic” residues are highly conserved in eukaryotic RdRPs including the single *Tetrahymena thermophila* cellular RdRP, Rdr1, where they were found to be required for RdRP activity *in vitro* (Collins & Lee, 2007). In contrast, in Rsp1, these residues appear to be absent from the predicted RdRP domain found in Rsp1, with the domain match ending just prior to where the catalytic residues are located in the RdRP domain, and were not found elsewhere in Rsp1 amino acid sequence. These findings suggest that although Rsp1 may have a protein domain that resembles an RdRP, it is unlikely to have catalytic activity of an RdRP.

**Identification of potential orthologs/homologs of Rsp1 in *Tetrahymena* species**

Next, I sought to determine if Rsp1 was a novel protein limited to *Tetrahymena thermophila* or if Rsp1 homologs exist in other organisms. To initially look among *Tetrahymena*, I performed Basic Local Alignment Search Tool (BLASTp and tBLASTn) searches in sequenced *Tetrahymena* species to identify potential Rsp1 homologs. These searches resulted in the discovery of several possible Rsp1 orthologs (Table 1). Notably, in *Tetrahymena thermophila*, the bonafide RdRP, Rdr1, came up as an Rsp1 hit, further underscroing the similarity between Rsp1 and RdRPs. Each BLAST search in other *Tetrahymena* species resulted in only one or two identified proteins, with the most significant hit annotated as “Hypothetical Rsp1” or “Hypothetical protein” and the less significant hit, if present, often annotated as “Hypothetical RdRP.”

The most significant hits had e-values ranging between 0 and 3.0e-31 and percent identities in aligned regions between 30.2% and 61.24%, with alignments to Rsp1 encompassing the
majority of their sequences (Table 1, red boxes). Two Tetrahymena species, T. pyriformus and T. canadensis, had a pair of highly truncated “Hypothetical Rsp1” hits which appeared to be neighbors in genomic location, suggesting that they may been incorrectly annotated as two separate genes in both species (Table 1). Reciprocal BLASTp searches of the T. thermophila genome using the amino acid sequences of the most significant hits returned Rsp1 as the top hit, confirming their relatedness. The second BLAST hit with Rsp1 in Tetrahymena species had e-values between 1.0x10^{-7} and 9.2x10^{-1} and percent identities that ranged from 18.4% to 24.5%. Reciprocal BLAST searches with these hits often returned the T. thermophila RdRP Rdr1 as the top hit. Further underscoring the sequence similarity between Rsp1 and RdRPs, the top scoring Rsp1 hits in T. empidokyrea and T. canadensis returned Rdr1 in addition to Rsp1 in reciprocal searches in T. thermophila. Similarly, the lower scoring Rsp1 hit in T. shanhaiensis returned both Rdr1 and Rsp1.

Given that Rsp1 contained a predicted RdRP domain, but lacked the catalytic residues, we were interested in assessing whether this was true for the Rsp1 hits identified above using NCBI CDD. This analysis revealed that all these proteins contain predicted RdRP domains, as shown in Table 1. However, consistent with the reciprocal BLAST searches described above, the RdRP domains of the most significant Rsp1 hits in Tetrahymena species appear to lack RdRP catalytic residues, just like Rsp1, whereas these residues are present in the less significant Rsp1 hits, with the following consensus sequence: “DLDGD” (Figure 3, top). Based on these analyses, the top scoring Rsp1 BLAST hits in other Tetrahymena species were termed “Rsp1-like”, while the lower scoring hits termed “Rdr1-like.” This delineation between Rsp1-like proteins and Rdr1-like proteins is visually noticeable in the percentage identity shading in the top portion of Figure 3. The darker shaded regions indicating higher percentage identity are present in grey bands among
the Rdr1-like proteins and dramatically reduced among the Rsp1-like proteins. Interestingly, when residues were colored based on their hydrophobicity and hydrophilicity, the common amino acid residue properties were highlighted across both the Rsp1-like and Rdr1-like protein groups (Figure 3, bottom). Phylogenetic analysis with MEGA-X of these Rsp1-like and Rdr1-like proteins in *Tetrahymena* species affirmed that the Rsp1 hits fall into two clear clades: Rsp1-like proteins and Rdr1-like RdRPs (Figure 4).

Interestingly, there was also one Rsp1 hit, the *T. malaccensis* Rdr1-like RdRP, where the sequence alignment with Rsp1 (red box, Table 1) overlapped the region containing the catalytic DxDxD residues found in canonical RdRPs. Moreover, the RdRP domain identified by NCBI CDD in the Rsp1-like proteins in *T. malaccensis*, *T. pyriformis*, and *T. canadensis* (green boxes, Table 1) encompassed the region that normally contains the catalytic residues in RdRPs, though, as discussed above, the catalytic residues appear to be missing. Rsp1 homologs and potentially related RdRPs like these may be key to understanding the evolutionary relationship between RdRPs and Rsp1-like proteins, as is further addressed in the discussion. Intriguingly, further BLAST analysis using *T. thermophila* Rdr1 performed on *Tetrahymena* species whose BLAST search with Rsp1 did not yield a hit to an RdRP (*T. elliotti*, *T. pyriformis*, *T. canadensis*, and *T. vorax*) revealed that these species appear to lack predicted RdRPs altogether. The significance of this is unclear but may suggest that Rsp1 plays a more important function in these organisms than an RdRP, though it is possible that an RdRP was missed due to missing sequence in incompletely sequenced genomes.

**Identification of potential orthologs/homologs of Rsp1 beyond *Tetrahymena***
Next, I expanded my search for homologs to other ciliates, unicellular eukaryotes, and other model organisms with sequenced genomes. We chose to first search for Rsp1 homologs in a number of ciliate species, namely, *Stentor coeruleus, Euplotes vannus, Oxytricha trifallax, Paramecium tetraurelia*, and *Stylonychia lemnae*, as these species have relatively well annotated genomes. Beyond these ciliates, we searched the unicellular *Dictyostelium, Giardia, Neurospora, Saccharomyces, Toxoplasma, Trypanosoma*, and multicellular *Arabidopsis, Ichthyophthirius, Caenorhabditis*, and several multi-species databases. These databases included species with well annotated genomes or species that are phylogenetically close to *Tetrahymena*, allowing us to gain a better understanding of the scope of the *RSP1* gene conservation across taxa. The top resulting protein hits from each of these searches is listed in Supplemental 1. Although not all searches yielded clear homologs, there were some hits that were found to be known RdRPs, proteins with a predicted RdRP domain as identified by NCBI CDD, or uncharacterized eukaryotic proteins that bear resemblance to bacterial DNA pol-III (Table 2). Notably, several of the species listed above have known or predicted RdRPs, although searches with Rsp1 only yielded an RdRP hit in *Oxytricha trifallax, Stentor coeruleus, Paramecium tetraurelia, Paramecium caudatum*, with e-values ranging from $1.0 \times 10^{-7}$ to 16 and percent identities from 19-25%. In these species, there are several predicted or experimentally confirmed RdRPs, yet searches with Rsp1 only yielded one or two of the RdRPs in each of these species. Reciprocal analysis only yielded an Rsp1 hit in *Stentor coeruleus* and *Paramecium tetraurelia*, possibly suggesting a more distant relationship between Rsp1 and the RdRP hits in *Oxytricha trifallax* and *Paramecium caudatum*. Intriguingly, the hit to *P. tetraurelia* RdRP Rdr4, as well as an uncharacterized protein in *Schistosoma rodhaini* with an RdRP domain as identified by NCBI CDD, is notable in that these two proteins apparently lack the conserved RdRP catalytic residues, like Rsp1.
Together, the analyses described above indicate that Rsp1 may bear some primary sequence resemblance to RdRPs, RdRP-domain containing proteins, and proteins with domains resembling bacterial DNA pol III in organisms beyond *Tetrahymena*. However, the strongest Rsp1 homologs appear to be limited to *Tetrahymena species*.

**Analysis of the structural homology between Rsp1 and RdRPs**

Despite the similarity between RdRPs and Rsp1 suggested from the above analysis, the primary sequence identity in aligned regions between RdRPs and Rsp1 is limited (Tables 1, 2). For example, *T. thermophila* Rdr1 only has 18.4% identity when compared to Rsp1 (Table 1). Therefore, to further assess the relationship between RdRPs and Rsp1, we sought to determine whether structural analysis algorithms would detect RdRPs when searching databases with a predicted structure of Rsp1. From this analysis, the only two cellular RdRPs to have had their structures solved (Qian et al., 2016; Salgado et al., 2006), *Neurospora crassa* QDE-1 and *Thielavia terrestris* QDE-1, were identified using HHpred and Phyre2. Several other RdRPs or hypothetical RdRPs were identified using HHblits (Table 3). The e-values for these structural hits ranged from 2.4x10^-6 to 92.

The structural search algorithm Phyre2 functions both by searching for related protein models and by generating a template-based model of the input sequence. Using this latter tool, the two Phyre2 QDE-1 template-based models of Rsp1 were downloaded and structurally assessed in PyMol (Figure 5, Supplemental 2). Based on these results, it appears that Rsp1 aligns more closely with *N. crassa* QDE-1 (100 atoms aligned, RMSD: 0.256) than *T. terrestris* QDE-1 (76 atoms aligned, RMSD: 0.199). In both cases, the Phyre2 modeled region of Rsp1 aligns with the “neck” region of the RdRP, located upstream of the catalytic residues of the RdRP, a region which
has no known function outside of its role in dimerization in QDE-1-type RdRPs (Figure 5) (Qian et al., 2016; Salgado et al., 2006). In both models, Rsp1 does not appear to have conserved residues to the QDE-1-aligned amino acid sequences, although the general helical shape and regions of hydrophobicity may be conserved (Supplemental 2), which could account for the structural homology of Rsp1 in this region with RdRPs despite lacking strong homology at the primary sequence level.

After assessment of the Phyre2 models, all putative Rsp1 homologs, orthologs, and a set of known RdRPs were aligned with the MEGA MUSCLE alignment tool and phylogenetically analyzed (Figures 6 and 7). In these alignments the hydrophobic and hydrophilic conserved regions between the Rsp1-like and Rdr1-like proteins was again noticeable (Figure 6, bottom). In agreement with the phylogenetic analysis performed with only Tetrahymena proteins (Figure 4), these results point towards two possible clades into which these protein sort: Rsp1-like proteins and Rdr1-like RdRPs (Figure 7).

Assessment of RdRC-dependent sRNA precursor transcript levels with and without Rsp1

Given the possible structural relationship between Rsp1 and RdRP proteins, we hypothesized that Rsp1 might play a role in an early step during the biogenesis of T. thermophila RdRC-dependent sRNAs, as the RdRP Rdr1 does, which might account for why Rsp1 is essential for sRNA accumulation. While a number of possibilities for Rsp1 function exist, we first tested whether Rsp1 might have a role in stabilizing precursor sRNA transcript levels. Without Rsp1, we predicted that the precursor transcripts would be susceptible to degradation, resulting in the observed loss of sRNAs.
In order to study precursor transcript levels with or without RSP1, we generated *T. thermophila* strains bearing a knockout (KO) in the *RSP1* gene in various strain backgrounds and used RT-PCR to confirm their genotype (Figure 8). Next, PCR primers for sRNA precursors PH/TASL2, PH/TASL3, PGC IB, and PGC IIIB, with actin as a control, were used on complementary DNA generated from parental or RSP1 KO strains. These different precursor transcripts were assessed because PGC IIIB, IB and PH/TASL transcripts are processed by different RNA-dependent RNA complexes (RdRCs) (Figure 2). As shown in Figure 9, however, only PGC IIIB was observed to change in levels upon RSP1 KO, and in fact, contrary to our prediction, the levels of PGC IIIB increased in the RSP1 KO cells rather than decreased. These findings indicate that Rsp1 is not needed for the stabilization of sRNA precursor transcripts (Figure 9). Thus, we next turned to examining a potential role for Rsp1 downstream of sRNA precursor accumulation in sRNA biogenesis.

**Analyzing a possible role of Rsp1 in RdRC component level regulation**

In the sRNA biogenesis pathway, the precursor sRNA transcripts are processed by a set of RNA dependent RNA complexes (RdRC) which contain Rdn1, Rdn2, Rdf1, Rdf2, and the central protein, Rdr1 (Talsky & Collins, 2012). Because RdRC-dependent sRNAs fail to accumulate in the absence of Rsp1, we hypothesized that RdRC component proteins require for Rsp1 for their stability. We focused our analysis on Rdr1, Rdn1 and Rdn2 because Rdr1 is a central component in all RdRCs, while Rdn1 and Rdn2 are found in separate RdRCs and therefore may be subject to differential regulation.

Through Western blotting cell lysates generated from vegetatively growing or starved *T. thermophila* strains expressing ZZ or 3x Flag-ZZ tagged RdRC components, we found that all
three proteins accumulate in RSP1 KOs to higher levels compared to strains in which Rsp1 remained present (Figure 10, Table 4, Supplemental 3). Quantitative Western blot analysis indicated fold-change increases in the knockout strains ranging from 3.13 to 10.32 for Rdn1, 7.15 to 16.74 for Rdn2 and 8.42 for Rdr1, depending on the loading control used. These results were contrary to our predictions for Rsp1-dependent stability of Rdr1, Rdn1 and Rdn2, which would have been a reduction, not an increase, in RdRC components levels in RSP1 KO cells. This result lead us to consider an alternative hypothesis for how Rsp1 might be important for sRNA biogenesis – that Rsp1 is involved in the proper assembly of RdRCs.

**RdRC composition analysis**

Although RdRC components appear to accumulate in the RSP1 knock out strains, it is possible that the observed lack of sRNAs in RSP1 knockouts could be explained by a malformation of the RdRC as a whole. We predicted, therefore, that Rsp1 may be playing a role in regulating the RdRC formation or stabilization. In order to assess the integrity of the RdRC, we compared RdRC composition in strains with and without RSP1 expressing tagged RdRC components using affinity purification and visualization of purified material on silver stained SDS-PAGE gels. With this technique, we predicted that we would be able to see any changes to RdRC composition, whether new proteins were bound or known proteins were missing.

Using antibodies against the ZZ tag on Rdn1 3xFZZ, Rdn2 3xFZZ, and Rdr1-ZZ, we were able to purify the RdRCs from strains with wildtype backgrounds and visualize known RdRC component bands, identified by their mobility on SDS-PAGE (Lee et al., 2009; Talsky & Collins, 2012). Additionally, in Rdr1-ZZ purifications, we noted a previously unreported faint protein band that migrates at a size consistent with Rsp1 (Figure 11). Moreover, though each of the three purifications were performed under high stringency conditions that were previously found to
destabilize Dcr2-RdRC interactions, a Dcr2-sized band was present in the Rdn2 and Rdr1 purifications (Figure 11, B and C). Additional analysis would be required to determine the identity of these Rsp1-sized and Dcr2-sized bands.

Remarkably, affinity purifications from RSP1 KO strains failed to recover Rdn2- and Rdn1-containing RdRCs detectable by silver staining from the RSP1 KO strains. To distinguish whether the RdRCs were not detectable because they simply failed to be eluted from purification beads or were indeed not purified in the first place, we performed Western blots on Rdn1 3xFZZ strain samples taken before and after the purification procedure comparing the wild-type and RSP1 knock out strains (Figure 12). Westerns revealed that no Rdn1-containing RdRCs were recovered in purifications, despite its presence in cells prior to preparation of extracts for purification. These results support a hypothesis that RdRCs are disrupted in the RSP1 KO strains. The exact role of Rsp1 in the RdRC formation, localization and/or stabilization remains unclear and will require future analysis, as proposed in the Discussion.
DISCUSSION

In this work, we sought to gain deeper understanding of the possible function that Rsp1 plays in the biogenesis of ~23-24 nt sRNAs in *Tetrahymena thermophila*. Previous work had established the requirement for Rsp1 for the accumulation of ~23-24 nt sRNAs that were known to be dependent on RdRCs, though a molecular understanding of this genetic requirement had yet to be elucidated. Through a combination of sequence analysis and experimentation, we have uncovered a potential evolutionary relationship between Rsp1 and cellular RNA-dependent RNA polymerases (RdRPs), as well as evidence that Rsp1 functions by regulating the RdRC in some physical way, possibly through stabilization, localization, and/or complex formation.

**Rsp1 may be related to RdRPs, but is unlikely to be catalytically active as an RdRP**

Our results indicate that Rsp1 has a protein domain with similarity to the RdRP domain found in cellular RdRPs but lacks the known catalytic residues of RdRPs. This homology between Rsp1 and RdRPs is limited, though detectable by conserved domain analysis with NCBI CDD and SMART, analysis of BLAST search results, structural prediction analysis, and in the hydrophobicity and hydrophilicity trends in MUSCLE alignment (*Tables 1, 2, 3, Figures 3, 5 and 6*). While it is possible that Rsp1 and Rsp1-like proteins may possess catalytic activity that depends on residues outside of the canonical catalytic residues in known RdRPs, efforts to test Rsp1 for RdRP activity *in vitro* have been negative (personal communication, Kathy Collins, UC-Berkeley), supporting the likelihood that Rsp1-like proteins are catalytically inactive as RdRPs. Moreover, the predicted structure of Rsp1 further supports the likelihood that Rsp1 lacks RdRP catalytic activity, given the notable lack of structural or sequence homology to the catalytic domain of RdRP QDE-1 (*Table 3, Figure 6, Supplemental 2*). Though the RdRP-like domain in Rsp1 aligned with
a handful of cellular RdRPs in both *Tetrahymena* and other species, it is notable that in organisms with more than one RdRP, only one or two were identified by BLAST searches (*Figures 6 and 7*), potentially suggesting that the Rsp1 RdRP-like domain is more similar to a subset of RdRPs.

Interestingly, the low bootstrap values calculated for each of our trees between the Rsp1-like and Rdr1-like clades supports our hypothesis that Rsp1-like and canonical RdRP genes have rapidly differentiated from each other (*Figures 4 and 7*). This is potentially due to at least two possibilities: 1) Rsp1-like proteins may be members of a rapidly diverging branch of cellular RdRPs that has lost the catalytic residues which are highly conserved in canonical RdRPs or 2) Rsp1-like proteins evolved separately from cellular RdRPs, derived from a common ancestor of RNA polymerases that lacked catalytic activity. Indeed, in support of the latter, examination of the evolutionary relationship between DNA-dependent RNA polymerases (DdRPs) and eukaryotic RNA-dependent RNA polymerases (RdRPs) has led to a proposal that the ancestor of both types of RNA polymerases only possessed RNA-binding capability without catalytic activity and served a cofactor for a ribozyme RNA polymerase (Iyer, Koonin, & Aravind, 2003). Along similar lines, it is notable that at least a few *Tetrahymena* species possess genes encoding Rsp1-like proteins but not RdRPs (see *Results*). There was also an intriguing alignment between Rsp1 and the Rdr1-like protein in *T. malaccensis* which suggested a region of homology between the predicted RdRP domain within Rsp1 and the predicted RdRP domain in *T. malaccensis* that was inclusive of the RdRP catalytic residues of the *T. malaccensis* Rdr1-like protein (*Table 1*). This result could suggest that Rsp1 derived from an RdRP-like protein then lost the catalytic residues. It may also suggest that Rsp1-like proteins and RdRPs evolved from a common ancestor that lacked catalytic residues. Recent research into the evolutionary history of RdRPs suggests that this protein domain may have first evolved without catalytic residues (Iyer et al., 2003). It is possible that these primordial
ancestors to modern RdRPs and DdRPs evolved into two protein branches; Rsp1-like proteins and the cellular RdRPs and DdRPs we know today. Intriguingly, some of the uncharacterized eukaryotic proteins returned by BLASTp outside of Tetrahymena species were found by NCBI CDD to contain domains related to the bacterial DNA polymerase DNA pol III. It is unclear if this result is due to homology that may exist between DNA-dependent DNA polymerases and RNA polymerases. However, this strange possible homology between Rsp1 and unconfirmed polymerase-type domains could be evidence of a vast phylogenetic difference between modern polymerases and the Rsp1-like proteins.

Generally speaking, rapid evolution of Rsp1-like genes is supported by our finding that even among the putative Rsp1 orthologs we identified in the Tetrahymena species, the sequence homology was modest, with percent identity values ranging from 30.1% to 61.24% (Table 1, Figure 3). In addition, evolution of Rsp1-like genes may be faster than canonical RdRP genes. Although we were able to find a few Rsp1-like proteins and RdRPs in species outside of Tetrahymena with limited homology to Rsp1, some of which are yet uncharacterized (Table 2, Supplemental 2), homologs to the T. thermophila RdRP Rdr1 were readily found in diverse eukaryotic species (Figure 7, data not shown). We speculate that there may be other Rsp1 homologs that are simply undetectable by BLAST searches with Rsp1 due to their high sequence divergence. Further analysis of different organisms’ sRNA biogenesis pathways may result in the discovery of other Rsp1-like proteins by their molecular or biochemical role in the pathway, rather than by sequence or structural homology. Interestingly, in Paramecium tetraurelia, there are two proteins with RdRP domains, RdR3 and RdR4, which, like Rsp1, appear to lack catalytic residues (Marker, Le Mouël, Meyer, & Simon, 2010). Although Rdr4 and not Rdr3 was found via a BLAST search with T. thermophila Rsp1, despite the greater apparent evolutionary relatedness of Rdr3 to
Rsp1 (Figure 7), it is possible both of these proteins may function in an sRNA biogenesis pathway in a manner that is analogous to Rsp1 in *Tetrahymena*.

We speculate that a relatively fast “clock” for sequence divergence among Rsp1-like proteins may be akin to the rapid divergence of amino acid sequence in viral proteins which permits greater flexibility in their target interactions (Pál, Papp, & Lercher, 2006). Although whether Rsp1-like proteins bind sRNA precursors or sRNA precursor-protein complexes remains to be established, high tolerance of amino acid substitutions may reflect a necessity for successful targeting of a wide variety and/or rapid evolution of sRNA precursor transcripts and/or their associated proteins for sRNA biogenesis.

Despite the amino acid sequence divergence in the RdRP domains of RdRP QDE-1 and *T. thermophila* Rsp1, the predicted structure of Rsp1 suggests that it may have a potentially similar structure to the RdRP neck region of QDE-1 (Figure 5, Supplemental 2). Although the function of the neck region of RdRPs has not been widely studied, analysis of *T. terresteris* QDE-1 has indicated that this protein region is at least responsible for homodimerization (Qian et al., 2016). It is possible that *T. thermophila* Rsp1 forms homodimers or potentially heterodimers with the RdRP Rdr1 via this protein domain. Although previous efforts to purify complexes containing Rdr1 and Rsp1 did not report evidence for heterodimerization between Rdr1 and Rsp1 (Lee & Collins, 2006; Lee et al., 2009; Talsky & Collins, 2012), close analysis of our silver stains of RdRCs purified using affinity-tagged Rdr1 revealed a doublet protein band at 124 kDa, one of which migrates at a size consistent with Rsp1 (Figure 11). Future experiments focused on directly determining whether Rdr1 and Rsp1 interact in a complex would help resolve this possibility, as would experiments examining a possible homodimerization of Rsp1 itself. Alternatively, it is possible that the RdRP neck-like domain in Rsp1 may be responsible for the binding interactions
between Rsp1 and Rdn1 since the RdRP Rdr1 also binds to Rdn1 (Lee et al., 2009; Talsky & Collins, 2012). Other proteins or nucleic acids that both Rdr1 and Rsp1 bind would also be good candidates for interacting with the RdRP neck-like region of Rsp1.

Conjectures aside, it is difficult to make any concrete predictions about the function of the RdRP domain in Rsp1 without future analysis. X-ray crystallography of Rsp1 and Rdr1 would greatly improve our understanding of both the function of Rsp1 and the structural relationship between these two proteins. Given that the only known structures of cellular RdRPs are in *N. crassa* (QDE-1) and *T. terresteris* (QDE-1), it would also advance our understanding of cellular RdRPs as a whole to assess the structures of both the *Tetrahymena* Rdr1 and Rsp1. In addition, site-directed mutagenesis of the region of Rsp1 that bears homology to RdRPs could uncover its biological and biochemical significance.

**Rsp1 regulates the RdRC**

Our findings suggest that Rsp1 appears not to stabilize, but rather repress RdRC component abundance. The Western blot results described above reveal a dramatic increase in Rdn1, Rdn2 and Rdr1 protein levels in the *RSP1* knockout strain (Table 4, Figure 10). Complementary to our results, unpublished RNAseq results collected by undergraduate Brian Miller of the Lee Lab indicate that these same RdRC components are in fact also repressed at the mRNA level in a manner dependent on Rsp1 (data not shown). Further analysis of the mechanism by which Rsp1 regulates these genes is required. However, interestingly, RNAseq data also seems to suggest that an *RDN2* knockout also results in an increase in *RDN1* and *RDR1* mRNA levels, suggesting that regulation of the RNAi genes may be responsive to the presence of a functioning RNAi pathway.
Given our Western blot and RNAseq data discussed above, Rsp1 does not appear to be required for sRNAs accumulation because of a stabilizing effect on RdRC components. Instead, from our failure to recover RdRCs in the absence of Rsp1 (Figures 11 and 12), despite the apparent overaccumulation of RdRC components in RSP1 knockout cells, we suggest three non-mutually exclusive hypotheses to explain the essentiality of Rsp1 for both sRNA accumulation and RdRC purification: Rsp1 regulates the normal composition, cell-extract stabilization or cellular localization of RdRCs. Any of these possibilities might account for the apparent requirement of RSP1 for a reduction in sRNA precursors like PGC IIIB (Figure 9), as the loss of RSP1 would be expected to disrupt normal processing to sRNAs.

If the composition or structure of RdRCs has changed in RSP1 knockouts, perhaps our inability to purify the RdRCs via a tagged component is due to structural changes in the protein itself or to formation of an aberrant complex which prevents the affinity tag from binding to purification beads. These possibilities seem less likely (though not impossible) given the identical results obtained for RdRC purifications performed independently with tagged Rdr1, Rdn2 or Rdn1. Alternatively, the stability of the RdRC in cell extracts may be decreased in the RSP1 knockout. This could result from a change in RdRC composition and/or structure, or an overall increase in the activity of cellular proteases in cell extracts. As a third possibility, RdRC localization could be disrupted by the RSP1 knockout, such that RdRCs may not remain soluble in extracts but pellet with other insoluble material during the ultracentrifugation step in the purification protocol.

To distinguish between these hypotheses, Western blots could be performed on cell extracts prepared for purifications either before or after the ultracentrifugation step to determine if RdRCs are in fact unstable in extracts, stable but insoluble, or stable and soluble, but unable to bind to purification beads. If the RdRCs are found to be unstable, it may be necessary to pursue alternative
preparations of cellular extracts for future purifications. If the RdRCs are found to be stable, but insoluble, microscopy experiments comparing the cellular localization of RdRCs in the presence or absence of Rsp1 may reveal Rsp1-dependent localization of RdRCs. Finally, if the RdRCs are found to be stable and soluble but appear to be unable to bind to purification beads, other biochemical characterization of the RdRCs may be necessary, including placing affinity tags in a different place on RdRC components (ie. N-termini of Rdn1 or Rdn2, instead of the current C-termini), gradient sedimentation analysis to examine changes in RdRC shape or size, and classic, non-affinity tag based, biochemical purifications of the RdRCs.

**Conclusions**

Regardless of the molecular mechanisms, it is clear from our results that the physical properties of RdRCs appear to be disrupted in the absence of Rsp1. Thus, we propose that Rsp1 functions in the sRNA biogenesis pathway by regulating RdRCs in some fashion, potentially explaining the previously reported loss of RdRC-dependent sRNAs in *RSP1* knockout strains. Moreover, our sequence analysis of Rsp1 suggests an evolutionary relationship between Rsp1 and RNA polymerases, such as the *T. thermophila* RdRP Rdr1. We found that Rsp1 has potential orthologs in *Tetrahymena* species and bears limited sequence and structural homology to RdRPs and uncharacterized proteins with a DNA polymerase domain. Our data suggests the existence of two closely related clades of RdRP-like proteins: those that are Rsp1-like which seem to lack the amino acid residues require for catalytically activity as RdRPs and those that are Rdr1-like, the majority of which possess the conserved catalytic residues of RdRPs. The Rsp1-like genes appear to be more rapidly diverging from one another and from the Rdr1-like RdRPs. This rapid divergence of genes remains an area of interest for future studies, as we do not know what biotic
factors are driving this evolution. We suggest it may be due to reduced selective pressures due to a requirement for binding flexibility to diverse or rapidly evolving targets.

* Tetrahymena thermophila* Rsp1 may be unique in its role as an RdRC regulator, as it is the first RdRC regulator ever to be described. However, the presence of Rsp1-like proteins in organisms which are known or predicted to contain RNAi biogenesis pathways may be evidence that other organisms contain a similar regulatory protein. This analysis of Rsp1 in the *T. thermophila* sRNA biogenesis pathway will aid in our understanding of the variety of mechanisms by which non-coding RNAs are processed and regulated. Given the importance of RNAi pathways in regulating a wide array of cellular processes, our research into the regulators of these pathways will contribute to our broader scientific understanding of molecular mechanisms that underlie these processes.
LITERATURE CITED


https://doi.org/10.1038/nprot.2012.085
https://doi.org/10.1038/nprot.2015.053
rna.061333.117. https://doi.org/10.1261/rna.061333.117
https://doi.org/10.1093/nar/gkx998
https://doi.org/10.1261/rna.1630309
https://doi.org/10.1093/nar/gky961
https://doi.org/10.1093/nar/gkq131
https://doi.org/10.1016/j.tig.2005.06.005


Yang, W., Jiang, C., Zhu, Y., Chen, K., Wang, G., Yuan, D., … Xiong, J. (2019). Tetrahymena...
https://doi.org/10.1093/database/baz029


Table 1. *Rsp1* orthologs/homologs in *Tetrahymena* species. Blastp searches were performed in *Tetrahymena* comparative genome database using the Rsp1 amino acid sequence. Predicted proteins designated “Hypothetical Rsp1” were the top hits in all cases; *T. thermophila* Rdr1 and “Hypothetical RdRP” proteins were the second best hit by e-value. % identity reflects frequency of residue identity across BLAST aligned sequence regions. Reciprocal BLASTp results indicate whether or not a *Tetrahymena* BLASTp with the Rsp1 orthologs/homologs resulted in an Rsp1 or Rdr1 hit. Red box: Region in hit that aligned to Rsp1, Green box: RdRP domain as predicted by NCBI CDD 3.16 unless otherwise noted, red line: RdRP catalytic residues: DxDxD (not to scale), *: RdRP domain hit by SMART Pfam 31.0, #: RdRP domain hit by SMART Pfam 32.0.

<table>
<thead>
<tr>
<th>RdRP Hit Models aligned to <em>T. thermophila</em> Rdr1(RdRP)</th>
<th>Species</th>
<th>Polymerase hits with aligned region**</th>
<th>e-values</th>
<th>% Identity</th>
<th>Reciprocal BlastP</th>
<th>e-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rsp1</td>
<td><em>T. thermophila</em></td>
<td></td>
<td>0.0E0</td>
<td>100</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Hypothetical Rsp1−</td>
<td><em>T. malaccensis</em></td>
<td></td>
<td>0.0E0</td>
<td>61.24</td>
<td>Rsp1</td>
<td>0</td>
</tr>
<tr>
<td>Hypothetical Rsp1</td>
<td><em>T. elliottii</em></td>
<td></td>
<td>0.0E0</td>
<td>50</td>
<td>Rsp1</td>
<td>0</td>
</tr>
<tr>
<td>Hypothetical Rsp1*</td>
<td><em>T. borealis</em></td>
<td></td>
<td>1.0E-130</td>
<td>37</td>
<td>Rsp1</td>
<td>1.00E-179</td>
</tr>
<tr>
<td>Hypothetical Rsp1</td>
<td><em>T. empidokryea</em></td>
<td></td>
<td>3.0E-111</td>
<td>30.2</td>
<td>Rsp1, Rdr1</td>
<td>e-126, 6e-14</td>
</tr>
<tr>
<td>Hypothetical Rsp1*</td>
<td><em>T. pyriformus</em></td>
<td></td>
<td>3.0E-31, 5.0E-84</td>
<td>30.1, 34.75</td>
<td>Rsp1</td>
<td>1.00E-91</td>
</tr>
<tr>
<td>Hypothetical Rsp1</td>
<td><em>T. canadensis</em></td>
<td></td>
<td>1.0E-56, 4.0E-74</td>
<td>38.25, 46.98</td>
<td>Rsp1, Rdr1</td>
<td>2e-67, 7e-6</td>
</tr>
<tr>
<td>Hypothetical Rsp1+</td>
<td><em>T. vorax</em></td>
<td></td>
<td>6.1E-66</td>
<td>35.8</td>
<td>Rsp1</td>
<td>6.00E-67</td>
</tr>
<tr>
<td>Hypothetical RdRP</td>
<td><em>T. borealis</em></td>
<td></td>
<td>1.0E-7</td>
<td>19</td>
<td>Rdr1</td>
<td>0</td>
</tr>
<tr>
<td>Hypothetical RdRP</td>
<td><em>T. malaccensis</em></td>
<td></td>
<td>2.2E-4</td>
<td>19.47</td>
<td>Rdr1</td>
<td>0</td>
</tr>
<tr>
<td>Rdr1</td>
<td><em>T. thermophila</em></td>
<td></td>
<td>1.2E-2</td>
<td>18.4</td>
<td>Rdr1, Rsp1</td>
<td>0, 8.4e-2</td>
</tr>
<tr>
<td>Hypothetical RdRP</td>
<td><em>T. empidokryea</em></td>
<td></td>
<td>2.9E-2</td>
<td>20.4</td>
<td>Rdr1</td>
<td>0</td>
</tr>
<tr>
<td>Hypothetical RdRP</td>
<td><em>T. shanghaiensis</em></td>
<td></td>
<td>4.6E-2</td>
<td>19.34</td>
<td>Rdr1, Rsp1</td>
<td>0, 8.0e-2</td>
</tr>
<tr>
<td>Hypothetical RdRP</td>
<td><em>T. paravorax</em></td>
<td></td>
<td>9.2E-1</td>
<td>24.5</td>
<td>Rdr1</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 2. Possible Rsp1 polymerase homologs in non-*Tetrahymena* species. Blastp searches were performed in a variety of databases. NCBI searches were restricted to the ciliates *Stentor, Oxytricha, and Paramecium*, respectively. Results from NCBI searches against *Tetrahymena* are included in the table for comparison. Color coding identical to Table 1 with the addition of a black bar indicating the DNA pol III predicted protein domain. Unc: unconfirmed protein, U.P.: UniprotKB, ParaDB: Paramecium Database, *O. trifallax* BLAST hit absent after 9/2017.

<table>
<thead>
<tr>
<th>RNA/DNA polymerase Hit</th>
<th>Species</th>
<th>Polymerase hits with aligned region**</th>
<th>e-values</th>
<th>% Identity</th>
<th>Reciprocal BlastP</th>
<th>e-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rsp1</td>
<td><em>T. thermophila</em></td>
<td></td>
<td>0.0E0</td>
<td>100</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>‘Rdq2’ (Rdr1)</td>
<td><em>T. thermophila</em></td>
<td></td>
<td>6.0E-3</td>
<td>18.4</td>
<td>Rsp1</td>
<td>1.2e-2</td>
</tr>
<tr>
<td>Rdq3</td>
<td><em>S. coeruleus</em></td>
<td></td>
<td>2.5E-2</td>
<td>20.33</td>
<td>Rsp1</td>
<td>2.3e-2</td>
</tr>
<tr>
<td>‘Rdq’ (Rdr1)*</td>
<td><em>O. trifallax</em></td>
<td></td>
<td>2.7E0</td>
<td>NA</td>
<td>No Hit</td>
<td>NA</td>
</tr>
<tr>
<td>‘Rdq’ (Rdr2)</td>
<td><em>P. tetraurelia</em></td>
<td></td>
<td>7.0E0</td>
<td>21.25</td>
<td>Rsp1</td>
<td>9.8e0</td>
</tr>
<tr>
<td>‘Hypothetical’ (Rdr4)</td>
<td><em>P. tetraurelia</em></td>
<td></td>
<td>1.6E+1</td>
<td>25</td>
<td>Rsp1</td>
<td>1.9e1</td>
</tr>
<tr>
<td>Rdq</td>
<td><em>P. caudatum</em></td>
<td></td>
<td>1.0E-7</td>
<td>19</td>
<td>No Hit</td>
<td>NA</td>
</tr>
<tr>
<td>Rdq2</td>
<td><em>P. tetraurelia</em></td>
<td></td>
<td>3.0E-6</td>
<td>19</td>
<td>Rsp1</td>
<td>9.8e0</td>
</tr>
<tr>
<td>‘Unc.’ (CDD: Rdq/Ddq)</td>
<td><em>S. rodhain</em></td>
<td></td>
<td>7.3E-9</td>
<td>20.7</td>
<td>No Hit</td>
<td>NA</td>
</tr>
<tr>
<td>Rdr1</td>
<td><em>T. thermophila</em></td>
<td></td>
<td>2.0E-6</td>
<td>18.1</td>
<td>Rsp1</td>
<td>1.2e-2</td>
</tr>
</tbody>
</table>

**DNA pol-III models aligned to *L. lactis* polC:**

| ‘Unc.’ (CDD: DNA pol-III) | *P. primaurelia* | 4.0E-11 | 20 | No Hit | NA |
| ‘Unc.’ (CDD: DNA pol-III) | *P. gallinaeum*  | 4.2E-7  | 20.6 | No Hit | NA |
| ‘Unc.’ (CDD: DNA pol-III) | *T. thermophila* | 5.1E-7  | 20.7 | No Hit | NA |
| ‘Unc.’ (CDD: DNA pol-III) | *N. californiae* | 5.7E-7  | 24.3 | No Hit | NA |
| ‘Unc.’ (CDD: DNA pol-III) | *T. thermophila* | 3.8E-6  | 20.2 | No Hit | NA |
| ‘Unc.’ (CDD: DNA pol-III) | *P. gallinaeum*  | 6.4E-6  | 20.8 | No Hit | NA |
| ‘Unc.’ (CDD: DNA pol-III) | *P. reticulum*   | 1.1E-5  | 22 | No Hit | NA |
| ‘Unc.’ (CDD: DNA pol-III) | *N. californiae* | 1.8E-5  | 21.7 | No Hit | NA |
Table 3. **Rsp1 predicted structural homology to RdRPs.** Searches were performed using a variety of structural modeling algorithms and databases. Color coding and Table columns are identical to Table 1.

<table>
<thead>
<tr>
<th>RdRP Hit</th>
<th>Species</th>
<th>Polymerase hits with aligned region**</th>
<th>confidence/e-values</th>
<th>% Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phyre2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QDE-1</td>
<td><em>N. crassa</em></td>
<td></td>
<td>85.3% confidence</td>
<td>18</td>
</tr>
<tr>
<td>QDE-1</td>
<td><em>T. terrestris</em></td>
<td></td>
<td>64.1% confidence</td>
<td>19</td>
</tr>
<tr>
<td><strong>SMART</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RdRP domain</td>
<td><em>Consensus</em></td>
<td></td>
<td>2.4E-6</td>
<td>-</td>
</tr>
<tr>
<td><strong>HHpred</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RdRP</td>
<td><em>Consensus</em></td>
<td></td>
<td>2.9E-2</td>
<td>16</td>
</tr>
<tr>
<td>QDE-1</td>
<td><em>T. terrestris</em></td>
<td></td>
<td>2.5E1</td>
<td>13</td>
</tr>
<tr>
<td><strong>CDD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QDE-1</td>
<td><em>N. crassa</em></td>
<td></td>
<td>4.4E1</td>
<td>19</td>
</tr>
<tr>
<td>QDE-1</td>
<td><em>N. crassa</em></td>
<td></td>
<td>6.3E1</td>
<td>20</td>
</tr>
<tr>
<td><strong>HHblits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RdRP domain</td>
<td><em>Consensus</em></td>
<td></td>
<td>9.2E1</td>
<td>17</td>
</tr>
<tr>
<td>RdRP-like 2</td>
<td><em>S. coeruleus</em></td>
<td></td>
<td>2.0E-2</td>
<td>-</td>
</tr>
<tr>
<td>Unc. RdRP</td>
<td><em>R. irregularis</em></td>
<td></td>
<td>4.9E-5</td>
<td>20</td>
</tr>
<tr>
<td>RdRP-like 3</td>
<td><em>S. coeruleus</em></td>
<td></td>
<td>5.9E-2</td>
<td>19</td>
</tr>
<tr>
<td>RdRP (CDD: RdRP)</td>
<td><em>P. parasitica</em></td>
<td></td>
<td>6.4E-1</td>
<td>22</td>
</tr>
<tr>
<td>Unc. (CDD: RdRP)</td>
<td><em>A. alpina</em></td>
<td></td>
<td>4.5E1</td>
<td>21</td>
</tr>
<tr>
<td>Unc. (CDD: RdRP)</td>
<td><em>C. sinensis</em></td>
<td></td>
<td>7.4E1</td>
<td>21</td>
</tr>
</tbody>
</table>

**Note:** Color coding indicates the level of homology.
Table 4. Quantification of RdRC component levels +/- Rsp1. Western blot bands were quantified using ImageStudioLite. Mean fold changes are for the normalized levels of the indicated protein in *RSP1* knockout strains divided by the levels in wildtype backgrounds. Signals were normalized to either α-tubulin or total protein (Ponceau) stains, which served as sample loading controls. P-value was obtained using one-sample two-tailed t-tests. N: number biological replicates.

<table>
<thead>
<tr>
<th></th>
<th>Ponceau Normalized</th>
<th></th>
<th></th>
<th>Tubulin Normalized</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Fold Change</td>
<td>P-value</td>
<td>N</td>
<td>Mean Fold Change</td>
<td>P-value</td>
<td>N</td>
</tr>
<tr>
<td>Rdn1</td>
<td>3.13</td>
<td>0.017</td>
<td>3</td>
<td>10.32</td>
<td>0.046</td>
<td>4</td>
</tr>
<tr>
<td>Rdn2</td>
<td>16.74</td>
<td>0.024</td>
<td>4</td>
<td>7.15</td>
<td>0.011</td>
<td>3</td>
</tr>
<tr>
<td>Rdr1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>8.42</td>
<td>0.012</td>
<td>3</td>
</tr>
</tbody>
</table>
Figure 3. *Tetrahymena* Rsp1-like proteins contain a domain with amino acid sequence similarity but limited sequence identity with *Tetrahymena* RdRPs. MEGA alignments generated by MUSCLE of the Rsp1-aligning region from all *Tetrahymena* Rsp1-like and RdRP-like proteins listed in Table 1, with the exception of sequences from *T. vorax* and *T. shanghaiensis*, which had a high frequency of large indels which prevented their alignment. The region containing catalytic DxDxD residues for RdRPs is boxed in black, with consensus sequences/properties shown either on top or on bottom of alignments and Rsp1 amino acid numbers at top (aa 1031-1140). **Top** Alignment depicting sequence identity. Darker shading reflects higher frequency of residue identity across aligned sequences. **Bottom** Alignment depicting conserved hydrophobicity (red)/hydrophilicity (blue) across aligned sequences.
Figure 4. Phylogenetic Analysis of putative *Tetrahymena* Rsp1-like and RdRP proteins. MEGA alignment of the full-length sequences of *Tetrahymena* Rsp1 orthologs and homologs, minus the *T. shanghaiensis* sequence due to large gaps in the alignment. Maximum Likelihood method and JTT matrix-based model with discrete gamma distribution (5 categories) used to model this tree, with branch lengths to scale. 0.5 scale bar indicated 50% genetic variation between branches of that scale size. Numeric code to the left of each sequence name refers to an NCBI or *Tetrahymena* Genome Database gene accession number.
Figure 5. Rsp1 has a domain with structural similarity to *Neurospora crassa* QDE1 but limited sequence homology. Phyre2 template-based model of Rsp1 against *N. crassa* QDE-1, 2j7n, aligned with *N. crassa* QDE-1 2j7o model (the identical structure). 100 atoms aligned in the “neck” region, RMSD of 0.256. A) QDE-1 shown in grey, Phyre2 predicted Rsp1 model in yellow. Red arrow depicts viewpoint shown in B). B) Close up of aligned RdRP “neck” region showing structural similarity. C) Close up of aligned RdRP “neck” region showing identical residues in red, similar residues in orange, and non-conserved in grey. Corresponding sequence alignment shown below. Rsp1 amino acid range: 701-772 aa.
Figure 6. *Tetrahymena* Rsp1-like proteins contain a domain with conserved amino acid sequence similarity but limited sequence identity with RdRPs beyond *Tetrahymena* species. Sequences selected are representative of the major branches from Figure 7. All Rdrl-like proteins within the same species as select Rsp1-like proteins were chosen to compare interspecies and intraspecies sequence similarity. Similar to Figure 3, the region containing catalytic DxDxD residues for RdRPs is boxed in black, with consensus sequences/properties shown either on top or on bottom of alignments. Rsp1 amino acids ranged from aa 1019-1122. Alignment regions with 90% sequence gaps were removed for ease of visualization. **Top and bottom:** MEGA alignment of Rsp1-like proteins and Rdrl-like with sequence identity or hydrophobicity/hydrophilicity, respectively, highlighted as in Figure 3.
Figure 7. Phylogenetic Analysis of *Tetrahymena* Rsp1-like proteins and potentially related RdRP proteins. Tree was generated using the parameters described for Figure 4 using a MEGA alignment of *T. thermophila* Rsp1, putative *Tetrahymena* Rsp1 homologs, and RdRPs identified through BlastP and structural homology searches with *T. thermophila* Rsp1 and other RdRPs found in those same species. Sequences from *A. alpina, C. sinensis, S. rodhaini, O. trifallax, T. vorax,* and *T. shanghaiensis* omitted due to large gaps in the alignment. Arrowheads mark sequences that were initially identified as BlastP or structural homology search hits with *T. thermophila* Rsp1. Red star indicates sequences that contain the DxDxD catalytic residues.
Figure 8. **RSP1 knockout strains confirmed with RT-PCR.** RT-PCR performed on mRNA targets indicated on the right of gels, with actin mRNA as an internal positive control for RNA integrity. Δ: RSP1 knockout; MW: DNA ladder; gDNA: genomic DNA; RT: reverse-transcriptase; WT: parental strain containing tagged protein without ΔRSP1.
Figure 9. siRNA precursor transcript levels +/- Rsp1. RT-PCR performed on the siRNA precursor transcripts indicated on the right of gels, with actin mRNA as an internal positive control for RNA integrity. Δ: RSP1 knockout; MW: DNA ladder; gDNA: genomic DNA; RT: reverse-transcriptase; WT: parental strain containing tagged protein without ΔRSP1.
Figure 10. RdRC component levels +/- Rsp1. Western blots were performed using antibodies against the ZZ tag on A) Rdn1, B) Rdn2, and C) Rdr1 on extracts from growing cells (Rdn1 3xFZZ and Rdr1-ZZ) or cells starved for 6 hrs (Rdn2 3xFZZ) containing or lacking Rsp1 (ΔRSP1). Expression on Western blots was normalized to α-tubulin (A, B and C, top) or total protein as stained by Ponceau S (A and B, bottom). Two-fold dilutions of each sample were used identify signals in the linear range for quantification. MW: protein ladder; WT: parental strain with no tagged protein; *: cross reacting band in 6 hour starved cells.
Figure 1. RdRC purifications +/- Rsp1. High stringency purifications were performed from overnight starved cultures. Bands corresponding in size to known RdRC components (and potential interacting partners) are denoted by the name of the component on the right side of the gel images. from tagged protein-expressing lines MW: protein ladder; WT: parental strain with no tagged protein; TEV: TEV protease, used for RdRC elution from purification beads; ☆: known RdRC component; *: possible proteolytic Rdr1 product generated during sample collection.
Figure 12. Western blot analysis of Rdn1 in protein samples from Rdn1 3xFZZ affinity purification. A) Boiled whole cell samples collected prior to the preparation of extracts for affinity purification of Rdn1-containing RdRCs. B) “IP” samples collected after RdRC affinity purification on IgG beads. “Beads” samples represent IgG beads after RdRC elution from beads with TEV protease. *: cross reacting band.