Spring 2023

An Overview of Background and Experimental Methods Used to Investigate a ~23-24 nucleotide small RNA pathway and its Links to Genome Stability

Abigail Ruddick

An Overview of Background and Experimental Methods Used to Investigate a ~23-24 nucleotide small RNA pathway and its Links to Genome Stability

Abigail Ruddick

Western Washington University

Honors Program Senior Capstone

June 6, 2023

An Introduction To RNA

Ribonucleic acid, commonly known as RNA, is the first product in gene expression and comes in multiple forms and types (Alberts, 2004). It is a nucleic acid that is similar to DNA but differs slightly at a chemical level and therefore performs different functions. The most well-known type of RNA is messenger RNA (mRNA) which is a prominent part of the "Central Dogma" – a process that contains the major steps of genetic information flow from DNA to proteins. Proteins carry out much of the work of a cell. The process of transcription generates mRNA, which is ultimately used as a template by the ribosome for synthesis of a polymer composed of amino acids called a polypeptide through a process called translation. This polypeptide chain will eventually fold into a functional protein with the aid of different factors that ensure proper folding.

Beyond mRNA, there are RNAs that do not code for protein and are called noncoding RNAs (ncRNAs). Some of these ncRNAs are involved in the steps of gene expression. For example, transfer RNA (tRNA) is used to match the correct amino acids with successive three nucleotide (nt) "coding" sequences (codons) in the mRNA during translation. Small nuclear RNAs help to process precursors to mRNA called pre-mRNAs. Small nucleolar RNAs (snoRNAs) work on processing and modifying ribosomal RNAs (rRNA), which are part of ribosomes and essential in the translation step of the Central Dogma. For polypeptide synthesis, rRNA functions as a ribozyme, which is an RNA molecule that also acts as an enzyme and catalyzes chemical reactions.

Relatively small ncRNAs ~18-25 nt in length are products of RNA interference (RNAi) pathways that are found in diverse eukaryotes (Gutbrod and Martienssen, 2020). There are multiple types of small RNAs, with the major classes being miRNAs, siRNAs, and piRNAs. MicroRNA (miRNA) is associated with post-transcriptional gene regulation. After a miRNA is produced from double-stranded RNA precursors though the action of the Dicer enzyme, it associates with proteins of the Argonaute family, creating an RNA-induced silencing complex (RISC). This complex associates with mRNAs that are complementary in sequence to the miRNA and Argonaute can cleave the targeted mRNA, effectively silencing gene expression. In some cases, no cleavage happens, but the mRNA is still silenced and then degraded through the action of other enzymes. A different RNAi system, employing small interfering RNA (siRNA) is used to protect cells against transposable elements or viruses. Again, Dicer cleaves double stranded RNA (dsRNA) coming into the cell from a virus, here into siRNAs. These siRNAs associate with Argonaute and the RISC complex. Transponson-derived or viral transcripts with complementary sequence to siRNAs are again targeted for degradation. Another system relies on piwi-interacting RNA (piRNA). These RNA molecules are made in cells and are used to block the movement of transposable elements. These RNAs couple with a Piwi protein instead of Argonaute and can either target transposon-derived RNAs for degradation or directly repress gene expression at the DNA level, inhibiting transcription.

Tetrahymena thermophila as a Model Organism

In *Tetrahymena thermophila,* a single celled eukaryote, the most well-studied RNAi pathway is involved in what could be considered an extreme form of silencing: DNA elimination. Like all ciliates, *Tetrahymena thermophila* contains two nuclei, a macronucleus (MAC) and a micronucleus (MIC) (Ruehle *et al.*, 2016). The MIC genome is diploid and contains 5 unique chromosomes. The MAC has far more chromosomes with 181 chromosomes, at about 90 copies each, and is the nucleus used to express genes (Zhou *et al*., 2022). When it comes to cell reproduction, *T. thermophila* has two options. During vegetative growth and division, the MIC divides mitotically and the MAC divides amitotically, with apparently random chromosome segregation. Vegetative growth is the typical mode of cell replication for *T. thermophila* under nutrient rich conditions. However, under starvation conditions, cells can undergo conjugation, where cells of different mating types will come together to undergo the processes of meiosis, gametogenesis, and gamete nucleus fusion in the MIC, forming the zygotic nucleus which contains MIC DNA content of both mating partners. This zygotic MIC-derived nucleus will be used to produce both the MICs and the MACs of progeny, while the parental MACs are destroyed.

It is during the process of new MAC development that takes place during new MAC development from a zygotic MIC during *T. thermophila* sexual reproduction that DNA elimination occurs (Kataoka and Mochizuki, 2011). Part of the process involves the fragmentation of 5 chromosomes in the MIC to 181 chromosomes in the MAC. At the same time, ~15% of the DNA found in the MIC is removed from the new MACs. These pieces of eliminated DNA are called Internal Eliminated Sequences (IESs), many of which are repetitive sequences that interestingly, are transposon-like. This removal is precise and guided by a method that prevents sequences in the parental MACs from being eliminated in the new MACs as they are developing while any sequences in the developing MAC that do not appear in the parental MAC is eliminated. This process uses one of the *Tetrahymena* specific Argonaute proteins called Twis, which are homologs of the Piwi proteins found in humans (Couvillion *et al*., 2009). Specifically, this pathway uses Twi1p, which associates with ~28-29 nt sRNAs called scan RNAs (scnRNAs) that are produced using one of three Dicer like proteins within *T. thermophila* called Dcl1. These scnRNAs guides Twi1p to matching IES sequences, where it promotes heterochromatin formation which eventually leads to IES elimination from the new MAC genome.

A ~23-24 nt sRNA Pathway in *Tetrahymena thermophila*

While the conjugation-specific DNA elimination RNAi pathway is well-understood in *Tetrahymena*, a second, less well-understood RNAi pathway also exists in *T. thermophila*. This pathway generates ~23-24 nt sRNAs throughout the life cycle of *Tetrahymena*, including the asexual growth and division phase, unlike the production of scanRNAs (Lee and Collins, 2006). In addition, the majority of the ~23-24 sRNAs match sequences found within the MAC genome, again unlike the scanRNAs. Further research on the sRNA biogenesis pathway showed the importance of three RNA dependent RNA polymerase complexes (RDRCs) in the biogenesis of ~23-24 nt sRNAs (Lee, Talsky, and Collins, 2009, Lee and Collins, 2006). RDRCs contain an enzyme from the RNA-templated RNA polymerase enzyme family, termed Rdr1 and one or two

of four Rdr1 associated proteins. The RDRCS catalyze the production of double-stranded RNA from single-stranded sRNA precursors. The double-stranded RNA intermediates are then processed by a Dicer homolog distinct from Dcl1 called Dcr2, a second Dicer homolog which is the most similar in protein domain structure to Dicers found in other eukaryotes. Two of the Rdr1-associated proteins are named Rdn1 and Rdn2 as they contain nucleotidyl transferase domains and biochemical activity of yet unknown biological function. The other two Rdr-1 associated proteins were named Rdf1 and Rdf2 and have not been found to bear recognizable protein domains. Curiously, Rdf1 and Rdf2 are encoded by genes located next to each other in the genome, with no other coding region in between them. This could imply that these genes arose through gene duplication, with both derived from the same ancestral gene initially. When the relative level of expression of the four genes was studied throughout different parts of the life cycle, it was found that they do not all peak in expression at the same time. *RDN2* mRNA expression peaks in conjugation, while the mRNA expression of *DCR2, RDN1, and RDR1* all peak during vegetative growth. When looking closer at the timeline of conjugation, both *RDN2* and *RDF1* mRNA expression peak in mid-conjugation. The mRNA expression of *RDF2* is high in vegetative growth, appears to peak in early conjugation, and is then downregulated by midconjugation.

The observations described above implied that the three known RDRCs have distinct functions, a suggestion that was then supported by fundings that different types of sRNA precursor transcripts are processed in a partially overlapping manner by RDRCs (Couvillion *et al*., 2009), For example, Rdn2 is responsible for sRNAs accumulating from predicted pseudogene clusters and Rdn2, Rdf1 and Rdf2 responsible for sRNAs from high copy repeat derived sRNAs and genomic loc predicted to express unusually structured RNAs. Pseudogenes are DNA sequences that appear like protein coding regions but lack regulatory regions needed to be expressed and/or function correctly, while high copy repeats are simple sequences that occur many times in a genome. After the ~23-24 nt sRNAs are created by RDRCs and Dcr2, they associate with Twis that are distinct from the Twis associated with conjugation-specific scanRNAs. Twis 2, 8 and 12 are highly expressed during asexual growth, with Twi7 also detectable under more sensitive assays. Of these four growth-expressed Twis, only Twi2, 7 and 8 associate predominantly with ~23-24 nt sRNAs. Interestingly, sequencing of sRNAs associated with Twi2 and Twi8 showed that the sRNAs they carry incompletely overlap in sequence. Twi2 was shown to associate with sRNAs derived from pseudo-gene clusters, loci predicted to express unusually structured RNAs, and high copy repeats. Twi8 has been shown to associate with sRNAs that are derived from protein coding genes, as well as sRNAs derived from-pseudo gene clusters and loci predicted to express unusually structured RNAs (Farley and Collins, 2017). These findings imply that like the RDRCs, at least Twi2 and Twi8 may have both overlapping and distinct functional specializations.

Link between the ~23-24 nt sRNA pathway and genome stability in *Tetrahymena thermophila*

Investigation of the biological function of the ~23-24 nt sRNA RNAi pathway revealed that the pathway may be important for genome integrity in *T. thermophila*. When studying cell

strains that had been mutated to have *RSP1* or *RDN2* knocked out, enlargement and increased frequency of extranuclear bodies called chromatin extrusion bodies (CEBs) was found (Lee *et al*., 2021). When gene expression in *RSP1*Δ and *RDN2*Δ were compared to the parental SB210 strain, genes found to be overexpressed were found to have predicted or known roles in homologous recombination, base excision repair (BER), DNA helicases, and kinases. As many of these pathways are involved in genome maintenance and repair, a potential link to genome stability was investigated by initially examining Rad51, a double stranded break (DSB) repair protein. Consistent with a role for Rsp1 and Rdn2 in genome integrity, Rad51 levels in *RSP1*Δ and *RDN2*Δ were found to be significantly elevated. In addition, an increase in MAC-localized Rad51 and γH2A.X, a second marker of double-stranded DNA breaks, was found in *RSP1*Δ and *RDN2*Δ. Knockout strains of one of the most highly growth-expressed sRNA-associating Twis, Twi8, exhibited increased levels of Rad51 and an increased amount of MAC localized Rad51 and γH2A.X. Together, these data suggest that the ~23-24 nt sRNAs produced in *T. thermophila* have an important role themselves in the maintenance of genome integrity.

Intriguingly, relatively recent work in diverse eukaryotes beyond *Tetrahymena* have also revealed RNAi pathway links to ensuring genome stability (Gutbrod and Martienssen, 2020). Several studies have shown the direct effects of removing these RNAi pathways. For example, when the Dicer protein is knocked down, the levels of γH2A.X increases in several different organisms (Peng and Karpen, 2009, Roche, Arcangioli, and Martienssen, 2016).Small RNAs (sRNAs) linked to DNA damage have been shown in *N. crassa*, *A*. *thaliana*, and *D. melanogaster*, among other organisms (Lee *et al*. 2009, Wei *et al*. 2012, Michalik, Bottcher, and Förstemann, 2012).

Outstanding questions about the genome protective RNAi pathway in *T. thermophila* Further investigating possible links between different Twis and RDRCs and genome instability is crucial, as it remains unclear whether only Rdn2-RDRCs and Twi8 are involved As noted above, Rdn2-RDRCs partially overlap with other RDRCs in their processing of certain sRNA precursors, while Twi8 associates with sRNAs that at least Twi2 is also known to associate with. In addition, Twi2 is encoded in a cluster of *TWI2*-like genes in the genome, making the other Twis encoded in this region an intriguing target for investigation. In order to investigate these outstanding questions, I aimed to both knockout the genes of interest out to examine the impact of these knockouts on levels of proteins that are markers for DNA damage, comparing the levels of the knockout with the parental strain SB210.

Using genetic design to knock proteins from the pathway out of the cell.

The process used to knock out genes is one that has been used in molecular and cellular biology labs for quite some time. This kind of genetic alteration takes advantage of an already existing process within the cell called homologous recombination (HR). HR occurs when cells recognize a DSB and use another copy of a chromosome to repair the damaged one. This means that if the flanking DNA sequence around a gene of interest is known, introducing a new piece of DNA with the same flanking sequence around it, the natural processes of the cell will cause the new piece of DNA to be taken up and replace the gene of interest. This is especially effective in *T. thermophila*, as HR is believed to be the main mechanism for DSB repair.

The whole process of making a *T. thermophila* gene knockout involves designing a circular piece of DNA, making many copies of it and then preparing it for introducing into *T. thermophila* in order to genetically alter the cells. In order to engineer and create copies of the DNA sequence needed to knockout the gene, a circular piece of DNA, called a plasmid, is designed that contains an antibiotic resistance gene flanked on either side by regions around the gene that is meant to be knocked out. The plasmid is inserted into the bacterium *E. coli* to grow up many copies. DNA is extracted from cells and then purified so only the plasmid DNA remains. Once the plasmid is purified, restriction enzymes are used to cut on either side of the flanks within the plasmid, making a piece of DNA that consists of the two flanks on either side with the antibiotic cassette in between. This piece of DNA is then purified and concentrated. Next, a gene gun is used to genetically alter the *T. thermophila* cells. Gold particles are coated with the cut DNA, and then shot at the cells at high velocity. After this, antibiotics are used to help select only for the cells that have the gene of interest knocked out. Cells are treated with higher and higher levels of the antibiotic that is countered in toxicity by the introduced antibiotic resistance gene. Increasing the concentration of antibiotic permits only the survival and division of cells with the greatest number of their chromosomes with the knocked-out gene. Due to the high chromosome copy number of *T. thermophila*, this step is very important, as the intent is to knock out every copy of the gene of interest from the MAC. Finally, the cells are put through a process to check that they are true knockouts. In order to do this, mRNA is collected from the cells, many copies of DNA are made from it, and then products are run on a gel to check for the presence of the gene of interest. If bands appear, then it is known that the gene has not been knocked out. For this process, mRNA has to be used, as the knockout process will not knock out the gene from the MIC, and as such using just DNA samples would be ineffective, as MIC DNA will show up for the gene of interest even in a true MAC knockout. After all of this has been done, the brand-new knockouts can be tested for levels of Rad51 to start to investigate the levels of DNA damage.

Measuring Rad51 levels in a mutant knockout strain relative to those in a control strain.

Western blotting is an important tool in molecular biology and is often able to give a snapshot of what is happening in an organism. It is used to look for the presence of proteins and then measure the relative amount of the protein. By using the specificity of antibodies against proteins, a protein of interest can be probed for out of a large quantity of protein. Often primary and secondary antibodies are used, with the primary antibody targeting the protein of interest and the secondary targeting the primary antibody. The secondary antibody often will have an enzyme attached to it, such as HRP, which produces light when a chemical developer is applied to it.

Western blotting starts by culturing a large population of cells and then bursting them open. The protein sample is run out on a gel, a process that separates the proteins based on weight. Afterwards, the protein is transferred to a nitrocellulose membrane. The membrane is then blocked with milk to prevent non-specific binding when antibodies are introduced. The membrane is incubated first in primary and then in secondary antibody. Next, a developer is used to be able to visualize the bands of the protein of interest. These bands are imaged and then analyzed for the levels of the protein of interest. By comparing the knockout strains to the

parental strain, it is possible to see if there has been an increase in Rad51, which would imply increased levels of DNA damage. This is extremely useful, as once new knockout strains are created, they can be tested for these proteins, and the extent to which they contribute to genome integrity can be examined.

Conclusion

Further work into studying this pathway is required to fully elucidate the link it has to genome stability. While the current study is in *T. thermophila*, there is potential that such a pathway may be evolutionarily conserved in other eukaryotes. Due to the newly discovered links to genome stability, there is potential that if this pathway is conserved in humans, deeper understanding of it could be used to understand, prevent, or even treat diseases like cancer that are related to genome instability. Overall, the importance of RNAi pathways cannot be overstated, and further research is required to truly understand all the roles these pathways play in eukaryotic organisms.

Citations

Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., & Walter, P. (2015). Molecular biology of the cell. Garland Science, Taylor and Francis Group.

Couvillion MT, Lee SR, Hogstad B, Malone CD, Tonkin LA, Sachidanandam R, Hannon GJ, Collins K (2009). Sequence, biogenesis, and function of diverse small RNA classes bound to the Piwi family proteins of *Tetrahymena thermophila.* Genes Dev 23, 2016-2032.

Farley BM, Collins K (2017). Trans-generational function of *Tetrahymena* Piwi protein Twi8p at distinctive non-coding RNA loci. RNA 23, 530–545.

Gutbrod MJ, Martienssen RA (2020). Conserved chromosomal functions of RNA interference. Nat Rev Genet 21, 311-331.

Kataoka K, Mochizuki K (2011). Programmed DNA elimination in *Tetrahymena*: a small RNA-mediated genome surveillance mechanism. Adv Exp Med Biol. 722, 156-173.

Lee, H.-C. *et al.* (2009). qiRNA is a new type of small interfering RNA induced by DNA damage. Nature459, 274–277.

Lee SR, Collins K (2006). Two classes of endogenous small RNAs in *Tetrahymena thermophila*. Genes Dev 20, 28–33.

Lee SR, Talsky KB, Collins K (2009). A single RNA-dependent RNA polymerase assembles with mutually exclusive nucleotidyl transferase subunits to direct different pathways of small RNA biogenesis. RNA 15, 1363–1374.

Lee SR, Pollard DA, Galati DF, Kelly ML, Miller B, Mong C, Morris MN, Roberts-Nygren K, Kapler GM, Zinkgraf M, Dang HQ, Branham E, Sasser J, Tessier E, Yoshiyama C, Matsumoto M, Turman G (2021). Disruption of a ~23-24 nucleotide small RNA pathway elevates DNA damage responses in Tetrahymena thermophila. Mol Biol Cell. 32(15), 1335-1346.

Michalik, K. M., Böttcher, R. & Förstemann, K (2012). A small RNA response at DNA ends in Drosophila. Nucleic Acids Res.40, 9596–9603.

Peng, J. C. & Karpen, G. H (2009). Heterochromatic genome stability requires regulators of histone H3 K9 methylation. PLoS Genet.5, e1000435.

Roche, B., Arcangioli, B. & Martienssen, R. A (2016). RNA interference is essential for cellular quiescence. Science354, aah5651.

Ruehle M, Orias E, Pearson C (2016). *Tetrahymena* as a unicellular model eukaryote: Genetic and genomic tools. Genetics 203(2), 649-665.

Wei, W. *et al*. (2012). A role for small RNAs in DNA double-strand break repair. Cell149, 101–112.

Zhou Y, Fu L, Mochizuki K, Xiong J, Miao W, Wang G (2022). Absolute quantification of chromosome copy numbers in the polyploid macronucleus of *Tetrahymena thermophila* at the single-cell level. J Eukaryot Microbiol. 69(4), e12907.