The Ping-Pong Cycle Fine Tunes piRNA Populations to Target Active Transposons

Brad Kaptur[1]

[1]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, 19104, USA

Abstract

Piwi-interacting RNA (piRNA) are small (23-29 bp), non-coding RNA molecules that form RNA-protein complexes and have been identified in varied organisms with different biogenesis pathways and modes of action. However, even with recent insights, there is a degree of uncertainty regarding their biogenesis and molecular function of piRNAs. In the Ping-Pong model of piRNA biogenesis, the piRNA generation process is coupled with RNA interference, which is mediated by three Piwi proteins known as PIWI, AUB, and AGO3. The available data on primary piRNA production is very limited and the proposed models even in the most recent literature are, as a result, highly speculative. Previous works have provided experimental evidence of the importance of the amplification loop in defense against transposons as underlined by studies of hybrid dysgenesis in *Drosophila.* In this work, preliminary modeling was done using a simplified model of the Piwi biogenesis and Ping-Pong pathways. A total of four models were created to preliminarily understand the biogenesis of Piwi-interacting RNAs from a stochastic simulation standpoint. Two of these models focused on a reaction scheme that involved Ping-Pong and attempted to characterize its behavior. One model suggested implications of a system in which one of the key complexes, AGO3, underwent competitive inhibition. The fourth model characterized a system that did not have the Ping-Pong cycle. Comparisons between these models revealed several insights, including three key points. First, the most important variable in the regulation of the Ping-Pong loop was the transcription rate of the *Max element* gene. Second, AGO3 may face competitive inhibition by the binding of parsed *cluster* mRNA, preventing binding to the *Max element* gene within the cluster loop. This competitive inhibition affects the number of Ping-Pong iterations as well as the loading of Piwi downstream. Third, a comparison between the germline and somatic-like model suggests that the Ping-Pong loop positively amplifies the initial signal and overall promotes Piwi loading when compared to the somatic-like control. While these models presented a preliminary understanding of the basics of a positive feedback or amplification loop roughly based on the dynamics of the Piwi system, there is a great deal of work to be done to gain a greater understanding of this system and a closer look at the dynamics of the Ping-Pong pathway.

1 Introduction

1.1 Piwi-Interacting RNAs

Piwi-interacting RNA (piRNA) are small (23-29 bp), non-coding RNA molecules that form RNA-protein complexes through interactions with Piwi proteins [1]. PiRNAs have basically no known defining sequence characteristics other than a strong propensity for a 5'-uridine and a weak bias toward an adenosine in their 10^{th} position [2].They have been identified in both vertebrates and invertebrates and have varying biogenesis pathways and modes of action. PiRNAs are unique from other small RNAs from the RNAi-based pathway in that they are not produced through double-stranded intermediates by the Dicer enzyme, but rather thought to be produced from long polycistronic RNA transcripts by a Dicer-independent mechanism in both mammals and *Drosophila* [2]. Though there is a degree of uncertainty regarding their biogenesis and molecular function, much progress has been made in the past several years in this area of research. One particular area of interest is the Ping-Pong pathway: The piRNA Ping-Pong pathway was first proposed from *Drosophila* studies where the piRNA would display a high frequency of sequence complementarity at the 5' ends, presumably as a result of some cyclic process. A leading hypothesis is that Aub and Ago3 interact to allow for cyclic refinement of piRNA [3].

1.2 Transposons

Transposons (TEs), sometimes known as "jumping genes," are mobile segments of DNA that can move around to different positions in the genome within a single cell. During this process they may cause mutations. They make up a large fraction of eukaryotic genomes, and approximately 5.3% of the fruit fly genomes [4]. Transposon silencing is especially important in the germline because it maintains the genetic information that is inherited by future generations [5]. In *Drosophila,* transposon insertion into genes has been linked to development of DDT resistance as well as resistance to organophosphate insecticides. However, most TE (transposable element) insertions appear to be deleterious to the host, and are selected against [4]. Approximately 120 TE families have been characterized in *D. melanogaster*. Several studies have provided evidence that piRNAs can repress activities of retrotransposons [4].

1.3 The Ping-Pong Model

In the Ping-Pong model, the piRNA generation process is coupled with RNA interference. The RNA interference is mediated by three Piwi proteins known as PIWI, AUB, and AGO3 [1]. In the Ping-Pong mechanism, primary piRNAs cleave sense transposon transcripts and simultaneously produce secondary piRNAs from the sense transposons that then cleave antisense transposon transcripts [2].This mechanism can be viewed as dependent on the transcription of both sense and anti-sense transposon transcripts. Alternatively, piRNAs may be produced through a double-stranded intermediate based on reports of an RNA-dependent RNA polymerase in *Drosophila* [2].The Ping-Pong mechanism is a form of a positive feedback loop, though it is unknown how the mechanism is started in the first place. The Ping-Pong model requires already existing—also known as "primary"—piRNAs, which are presumably derived from clusters, to initiate the amplification cycle [5]. Mutations that eliminate primary piRNAs are predicted, by current models, to lead to the collapse of the Ping-Pong cycle [5]. Mutations in squash and zucchini (encoders of putative nucleases that localize to the perinuclear nuage), reduce piRNA levels without blocking Ping-Pong bias [5]. A similar mechanism is found in *Arabidopsis* for a different class of small RNAs, which suggests that mechanisms of this form may be universal. More specifically, mechanisms by which transposons are activated outside of the germline to generate small RNAs, thereby reducing the chance of deleterious transposon insertions may be prevalent in multiple forms. The Ping-Pong cycle requires initiators to amplify piRNAs; piRNA populations maternally inherited by germline transmission serve as important initiators of the Ping-Pong cycle to amplify piRNAs in the ovaries [3].The available data on primary piRNA production is very limited and the proposed models even in the most recent literature are, as a result, highly speculative [5].

1.4 piRNA Pathways in Somatic and Germline Cells

There are substantial biochemical differences between the piRNA pathways of somatic and germline cells. These differences are partially characterized in the model organism *Drosophila*. In the somatic piRNA pathway, piRNA cluster transcripts are parsed into shorter fragments that are loaded onto Piwi. At this step, Piwi likely selects precursors with a 5' uridine and the 3' tail of Piwi-bound RNAs are subsequently trimmed and the 2'-OH is methylated to generated mature piRNAs [6]. At this point, the mature piRNA complex silences various gene targets. In the somatic mechanism, several gene targets have been identified such as mdg1, stalker4, blood, roo, and gypsy, among others [7].

In germline cells, Ping-Pong amplification is used in addition to the primary piRNA pathway. During this biogenesis process, cluster transcripts generate sense and antisense piRNAs. The long piRNA precursor transcripts are parsed by unknown mechanisms into shorter fragments that are tails of pre-piRNAs, which are trimmed and methylated at the 2'OH to generate mature piRNAs. At this point, piRISCs with antisense piRNAs are sufficient to silence sequence complementary transcripts of active transposable elements. Separately, an active copy of a *Max* element is transcribed, which can be silenced by a complementary piRISC. The Ping-Pong cycle takes in loaded Aub and Piwi complexes as well as the transcribed *Max* element [7]. Initially, the Aub complex that is antisense to an active sense *Max* transcript recognizes the sequence, which guides slicing and loading onto AGO3. That complex performs two functions: First, it cleaves complementary cluster transcripts. Second, it initiates production of an antisense piRNA (the sequence of which is identical to the initiating piRNA), which may be either loaded onto Piwi, which would allow for repression of active transposons, or it may be loaded onto Aub, which would effectively continue the cycle, since the loaded Aub is identical to the aforementioned Aub complex that initiated the slicer cleavage [7]. As an additional note, maternal piRNAs may serve important roles in the starting phase of the Ping-Pong cycle, but these are currently unclear and thus will be ignored in the simulations presented in this work.

1.5 Refinement Hypothesis

In the outline of piRNA pathways in germline cells presented above, the Ping-Pong cycle as described would be a form of a positive feedback loop. Clearly there are differences between the piRNA biogenesis in somatic and germline cells, but the absence or presence of the Ping-Pong pathway has not been fully explained. Previous works have provided experimental evidence of the importance of the amplification loop in defense against transposons as underlined by studies of hybrid dysgenesis in *Drosophila* [4]. This work will use stochastic simulation to test the hypothesis that the Ping-Pong cycle fine tunes the piRNA population to only target active transposable elements, building upon those hypotheses as well as previous simulation work regarding piRNA [4,8].

2 Methods

2.1 Stochastic Gillespie Overview

These simulations as a whole use a stochastic Gillespie procedure [9] to simulate a simplification of the Ping-Pong pathway—and preceding steps involved in transcription and subsequent steps in TE transcript silencing. The following assumptions were inherent in these simulations:

(1) 3' end-trimming in all cases was instantaneous and not a rate-limiting step. (2) The maternal piRNA pool played no role in either the initial conditions or outcome of the Ping-Pong cycle. (3) Only parsed transcripts that targeted active TEs were recorded; parsed transcripts that did not target active TEs were ignored and were assumed not to negatively affect loading via competition or the simulation by any other unknown mechanism. (4) Loading was irreversible; in conjunction with assumption 3, this would be unlikely to affect the results of the simulation and made computation simpler. (5) Only one gene is present and it is always active. While this would not be true *in vivo*, it simplifies the simulation to a case

where one is only interested in the Ping-Pong cycle at short timescales while the gene is active and one is only interested in one gene in one cell. (6) No additional Piwi, AGO3, or Aub were produced during the simulation. While this is not necessarily true, this assumption simplified the simulation. Additionally, factoring in their transcription would not inherently change the outcome of the simulation unless their transcription was rate limiting some downstream process and would unnecessarily complicate the simulation. (7) The decision of whether Piwi or Aub would be loaded after loaded AGO3-mediated slicer cleavage is due to the relative abundances of unloaded Piwi and Aub. That is, there is no inherent regulatory mechanism that would result in a preference toward continuing the amplification loop or outputting loaded Piwi given the limited resource of only one transcript. (8) Slicer ends that were trimmed were immediately discarded. (9) Bidirectional transcription in the case of the *cluster* gene was performed in one step. (10) Loaded Piwi is the only complex capable of silencing active TE transcripts. Similarly, for a confined cell simulation the number of TE transcripts silenced scales proportionally with the number of loaded Piwi complexes active in the cell, and thus does not need to be detailed in the body of the simulation. (11) Transcription is rate-limited by the binding event and thus gene size does not affect the rate of transcription. (12) The events in the model are not spatially separated (thus, there is no time delay due to translocation). (13) The subset of parsed cluster RNA that targeted active TE transcripts was effectively identical to the unloaded product RNA of the loaded AGO3-mediated slicer cleavage in the Ping-Pong cycle, both of which will be interchangeably referred to from this point as parsed RNA. (14) The unloaded product RNA of the loaded Aub-mediated slicer cleavage in the Ping-Pong cycle only reacts with unbound AGO3 to form loaded AGO3 and undergoes no other reactions or takes part in any inhibitory mechanisms.

2.2 Model A: Germline piRNA Pathway

These simulations used the germline piRNA pathway as described in Senti, 2010 for the basis of the model [7]. Specifically, the reactions used are described in **Figure 1.**

$$(1)\ gene \xrightarrow{\mu_{cluster}} cluster$$
$$(2)\ gene \xrightarrow{\mu_{element}} active\ element$$
$$(3)\ cluster \xrightarrow{k_{parsed}} parsed\ RNA$$
$$(4)\ cluster \xrightarrow{\delta_{cluster}} \emptyset$$
$$(5)\ active\ element \xrightarrow{\delta_{cluster}} \emptyset$$
$$(6)\ parsed\ RNA \xrightarrow{\delta_{parsed}} \emptyset$$
$$(7)\ Aub + parsed\ RNA \xrightarrow{k_{Aub}} loaded\ Aub$$
$$(8)\ Piwi + parsed\ RNA \xrightarrow{k_{Piwi}} loaded\ Piwi$$
$$(9)\ loaded\ Aub + active\ element \xrightarrow{k_{slicer,Aub}} loaded\ Aub + cleaved\ RNA$$
$$(10)\ cleaved\ RNA + AGO3 \xrightarrow{k_{AGO3}} loaded\ AGO3$$
$$(11)\ loaded\ AGO3 + cluster \xrightarrow{k_{slicer,AGO3}} loaded\ AGO3 + \ parsed\ RNA$$
$$(12)\ cleaved\ RNA \xrightarrow{\delta_{cleaved}} \emptyset$$

Figure 1: Reaction diagram detailing reactions involved in germline piRNA pathway.

In addition to these reaction conditions and the aforementioned simplifications, it was necessary to make several other decisions, including length of duration, initial parameter amounts, and rate constant values. Many of the parameters had not been measured—at least in drosophila, so homologues in other organisms were used in several cases. In other cases, estimates were not available, so these were tested in different scenarios. For mRNA degradation, the measured parameters in the well-characterized organism *E. coli*

were used, and those parameters were taken from Thattai et al. and incorporated into the simulation [10]. As initialization conditions, the number of transcribed clusters, max elements, parsed mRNA, cleaved mRNA, loaded AGO3, loaded Aub, and loaded Piwi were all assumed to be 0. The *E. coli* parameters for transcription and degradation were applied to the transcription and degradation respectively of the *cluster* mRNA, the *max element* mRNA, as well as the degradation of the parsed and cleaved mRNA. Degradation was set to 0.006, and initiation to 0.003 [10]. Remaining parameters with no available measurement or estimate in the literature include the initial quantities of AGO3, Piwi, and Aub, as well as the rate constants for parsing, loading of AGO3, Piwi, and Aub, and the number of RNA that would target active transposons from the parsing of one *cluster* gene. For all programs, 1,000 simulations were used and each lasted 10,000 seconds.

Once this model was established, observing the parameter changes qualitatively was of interest for refining the model initially and gaining greater insight. This was performed by varying the non-fixed parameters in the simulations and recording qualitative shifts made to the output of the simulation.

2.3 Model B: Germline piRNA Pathway—Ping-Pong Intensive

After this initial assessment, the parameters were varied within physiologically reasonable constraints to generate a model at which the Ping-Pong Pathway would have the most effect using the same setup as in Model A, but with substantially different parameters. This model is referred to as Model B and reflects a variant of Model A where the effect of Ping-Pong is most evident—but still fits biological constraints.

2.4 Model C: Germline piRNA Pathway—Ping-Pong Intensive, AGO3 Competition

One step that was not clear in the paper by Senti et all was the role of loading AGO3 with a parsed *cluster* transcript with a 5' end that could start with any base [7]. Unlike Aub and Piwi which have clearly defined roles in the Ping-Pong cycle after being loaded with a parsed mRNA, the role of loading AGO3 with a *cluster* mRNA is unclear. Here, it is hypothesized that this would inhibit the loading of AGO3 with the cleaved *Max element* mRNA. An additional reaction is suggested in **Figure 2** for this model.

$$(10)\ cleaved\ RNA + AGO3 \xrightarrow{k_{AGO3,cleaved,}} loaded\ AGO3$$

$$(13)\ parsed\ RNA + AGO3 \xrightarrow{k_{AGO3,parsed}} blocked\ AGO3$$

Figure 2: Additional and modified equations corresponding to Model C, specifically regarding the hypothesized competitive inhibition of AGO3. This builds on the equations in Figure 1, with the exception of equation (10), which is redefined here.

2.5 Model D: Somatic Pathway (No Ping-Pong)

To ascertain the true effect of the Ping-Pong cycle on the *drosophila* germline pathway, a comparison was made to a generated model that mimicked the somatic pathway in *drosophila*. This model was generated using again the Model A equation scheme as a baseline, with mostly deletions. The equation list for this model is detailed in **Figure 3.**

$$(1)\ gene \xrightarrow{\mu_{cluster}} cluster$$
$$(2)\ cluster \xrightarrow{k_{parsed}} parsed\ RNA$$
$$(3)\ cluster \xrightarrow{\delta_{cluster}} \emptyset$$
$$(4)\ parsed\ RNA \xrightarrow{\delta_{parsed}} \emptyset$$
$$(5)\ Piwi + parsed\ RNA \xrightarrow{k_{Piwi}} loaded\ Piwi$$

Figure 3: Reaction scheme for the model of the somatic pathway involving no Ping-Pong amplification loop. In this model, the key components are the gene, the transcribed mRNA, the parsed cluster, and the Piwi protein (and its loaded form).

3 Results

3.1 Model A: Germline piRNA Pathway

In addition to the parameters previously set for initiation and degradation, the following parameters were chosen. This initial model puts the loading rate constants of Piwi, Aub, Ago3 and the slicer activity for loaded Aub and AGO3 as equivalent and equal to 0.001. Additionally, the parse rate is equal to the rate of transcription. It is assumed that 1 active mRNA is created per parsed gene and that there are 10,000 units of AGO3, Aub, and Piwi in a cell. The results of this simulation are outputted in **Figure 4** as histograms.

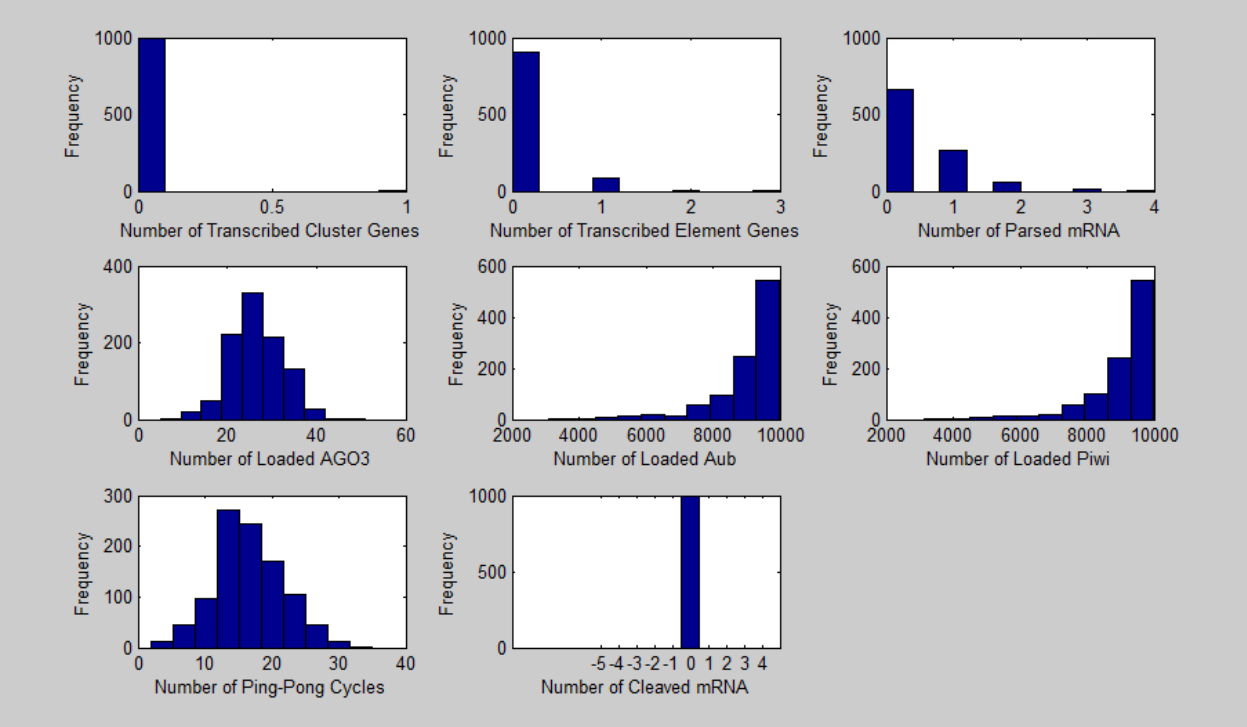

Figure 4: Outputted parameters of interest using the preliminary simulation parameters. Of interest were the number of existing genes of each type at the end of the simulation, the number of mRNA of each type, and the number of loaded complexes of each of the Piwi proteins. Of additional interest was a parameter that represented how many cycles of Ping-Pong were undergone in the simulation.

According to the results displayed in **Figure 4** for the preliminary parameters, the number of ping pong cycles was roughly normally distributed with mean 26.8 and standard deviation 5.2. At the end of each simulation, there were no cleaved mRNA present and only a small number of simulations had any transcribed cluster genes present. The number of transcribed element genes followed a Poisson distribution with mean 0.1010 and variance 0.1089. The number of parsed mRNA appeared to follow a Poisson distribution as well with mean 0.4200 and variance 0.4360. The number of loaded AGO3 appeared normally distributed with mean 26.7 and variance 5.9. The number of loaded Aub and loaded Piwi were each skewed right and appeared to follow similar distributions with mean of variance of 9047 and 9.999E5 and 9047 and 1.002E6 respectively.

3.2 Model A: Germline piRNA Pathway Parameter Modifications

Modifications were made successively to the parameters in the simulation to observe how parameter changes affected the results of the simulation. A holistic assessment is described in **Figure 5**.

Modification	Observed Changes
Num_parsed = 10; (10-fold increase)	Mean number of loaded Aub, loaded Piwi shifted right
Num_parsed = 0;	Mean Number of Parsed mRNA, loaded AGO3, loaded Aub, loaded Piwi, Ping-Pong cycles, and cleaved mRNA all go to 0
k_load_AGO3 = 1; **k_load_Aub = 1;** **k_load_Piwi = 1;** **(1000-fold increases)**	Number of Transcribed Cluster Genes goes to 0; Number of loaded Aub and loaded Piwi goes to 10,000
k_slicer_Aub = 1; **k_slicer_AGO3 = 1;** **(1000-fold increases)**	Number of Transcribed Element Genes goes to 0;
k_load_AGO3 = 1; **k_load_Aub = 1;** **k_load_Piwi = 1;** **k_slicer_Aub = 1;** **k_slicer_AGO3 = 1;** **(1000-fold increases)**	Mean Number of Loaded Aub, loaded Piwi increase
k_parse = 3 (1000-fold increase)	Mean number of parsed mRNA increases
mu_elem=0.03 (10-fold increase)	Mean number of parsed mRNA, transcribed element genes, number of Ping-Pong cycles, Loaded Aub, Loaded Piwi increase

Figure 5: Parameter modifications made to the code, which is attached as a Supplementary File. Most changes only made slight effects to the simulation, even at 1000-fold increases. However, changing the parameter mu_elem, which corresponds to the transcription rate for the *Max element* gene drastically increased the number of Ping-Pong cycles undergone as well as the number of loaded Aub and Piwi, which suggests that this parameter was limiting the Ping-Pong amplification cycle.

3.3 Model B: Germline piRNA Pathway—Ping-Pong Intensive

Compared to Model A, a drastic increase in the number of Ping-Pong cycles was achieved with a marked increase in the parameter corresponding to the transcription of the *max element* gene. While this initiation

rate (a 1000-fold increase from Model A), is likely outside the realm of biological possibility, it could be accounted for by either multiple genes, or a relatively higher degree of polymerase activity than then *cluster* gene. The results of this model are shown graphically in **Figure 6.**

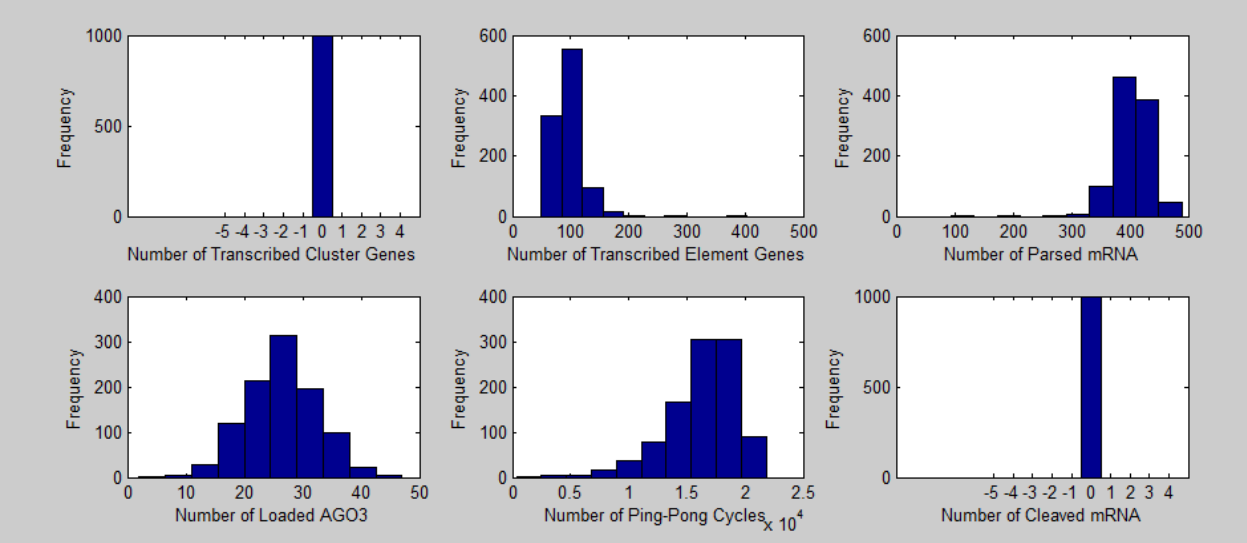

Figure 6: Outputted parameters of interest using the simulation parameters increasing Ping-Pong cycles. Of interest were the number of existing genes of each type at the end of the simulation, the number of mRNA of each type, and the number of loaded complexes of AGO3. Of additional interest was a parameter that represented how many cycles of Ping-Pong were undergone in the simulation. Of additional note (not pictured) was that in all 1000 simulations, all 10,000 of Piwi and Aub were loaded.

3.4 Model C: Germline piRNA Pathway—Ping-Pong Intensive, AGO3 Competition

Preliminarily, the same parameters as in Model A were used in the simulations for Model C. The new parameter that was reflective of the rate constant for loading the mRNA from the parsed *cluster* into AGO3 was the key differentiator between the two models. Initially, the parameter was set as equal to that of AGO3 loading for the sliced *Max element*, which had a value of 0.001. This yielded the output shown in **Figure 7.**

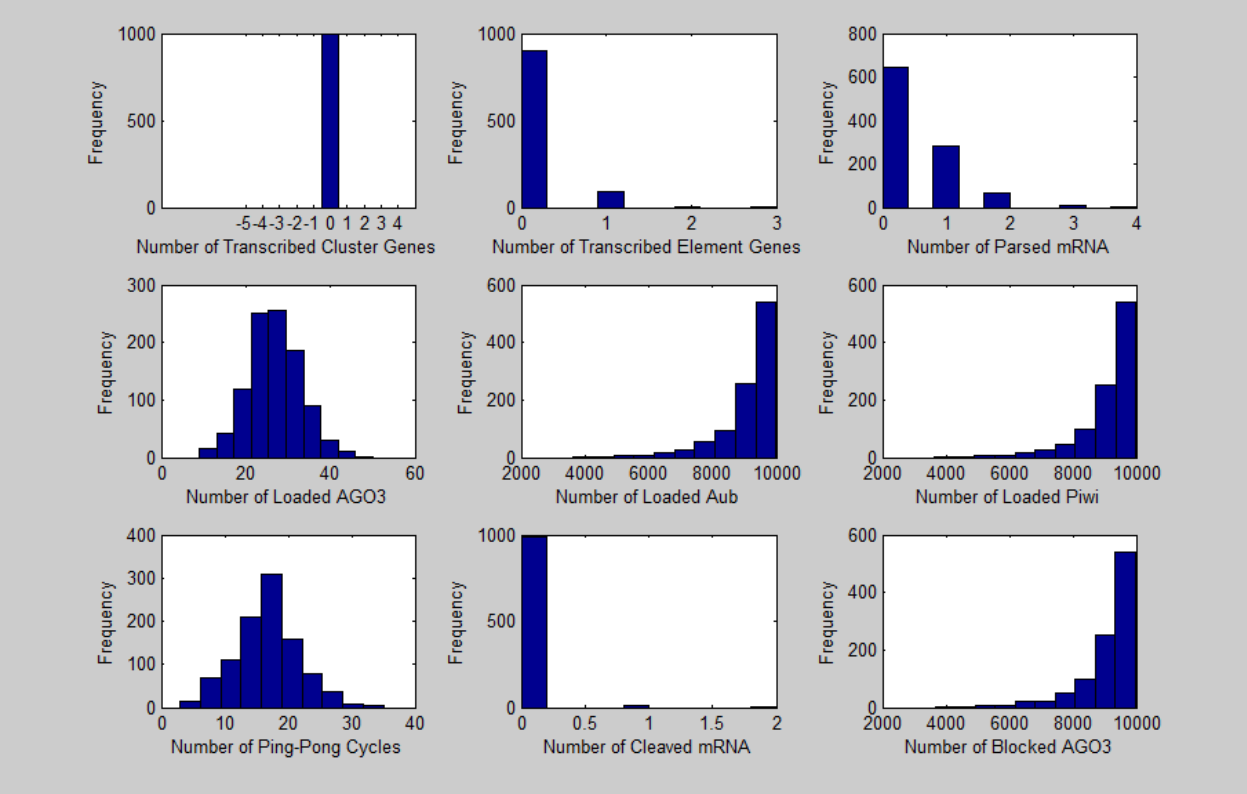

Figure 7: Outputted parameters of interest using the preliminary simulation parameters. Of interest were the number of existing genes of each type at the end of the simulation, the number of mRNA of each type, and the number of loaded complexes of each of the Piwi proteins. Of additional interest was a parameter that represented how many cycles of Ping-Pong were undergone in the simulation. In this simulation, the new parameter representing the number of blocked AGO3 was also displayed. As evident in this output, using the setup of having the blocking constant equivalent to the loading rate constant of AGO3 resulted in a drastically higher proportion of blocked AGO3, but did not substantially affect the number of Ping-Pong cycles or the number of loaded AGO3.

Additionally, the rate constant for blocking AGO3 was both increased by a factor of 1000 (**Figure 8)** and decreased by a factor of 1000 (**Figure 9**).

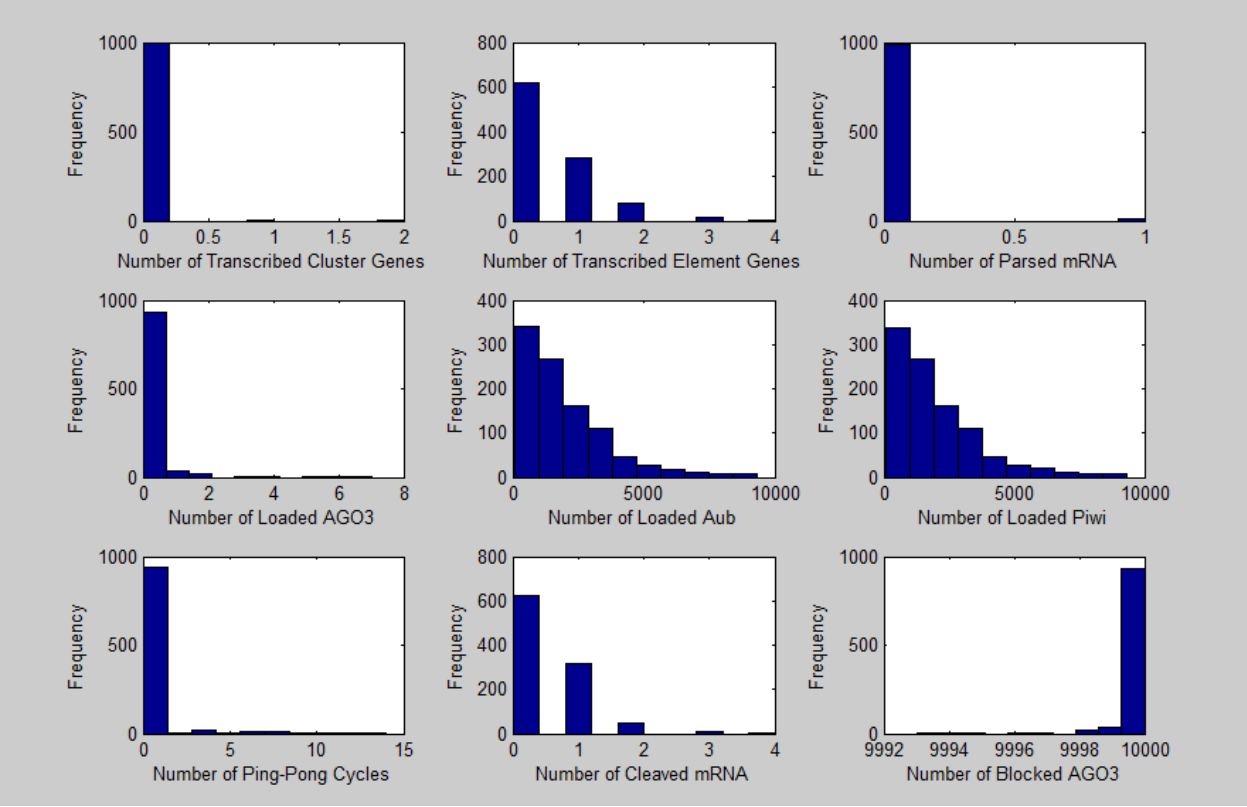

Figure 8: Outputted parameters of interest using the preliminary simulation parameters. Of interest were the number of existing genes of each type at the end of the simulation, the number of mRNA of each type, and the number of loaded complexes of each of the Piwi proteins. Of additional interest was a parameter that represented how many cycles of Ping-Pong were undergone in the simulation. In this simulation, the new parameter representing the number of blocked AGO3 was also displayed. As evident in this output, using the setup of having the blocking constant much greater (1000x) than the loading rate constant of AGO3 resulted in drastic changes to the simulation, affecting nearly all parameters compared to **Figure 6.**

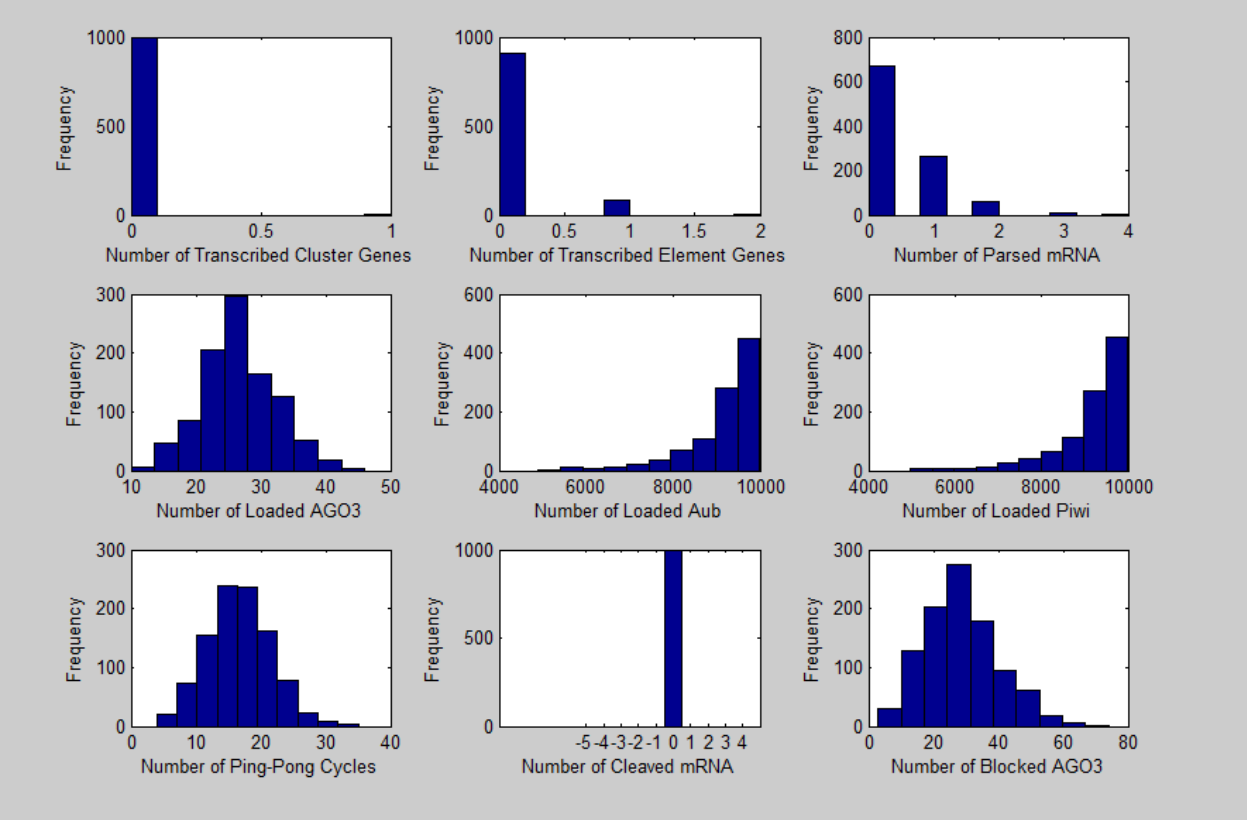

Figure 9: Outputted parameters of interest using the preliminary simulation parameters. Of interest were the number of existing genes of each type at the end of the simulation, the number of mRNA of each type, and the number of loaded complexes of each of the Piwi proteins. Of additional interest was a parameter that represented how many cycles of Ping-Pong were undergone in the simulation. In this simulation, the new parameter representing the number of blocked AGO3 was also displayed. As evident in this output, using the setup of having the blocking constant much less than (1/1000) the loading rate constant of AGO3 resulted in minor changes to the simulation, affecting only the number of blocked AGO3 compared to **Figure 6.**

3.5 Model D: Somatic Pathway (No Ping-Pong)

Model D is a variant of Model A that is reflective of the somatic (non-Ping Pong) pathway in *Drosophila* rather than the germline (involving-Ping Pong) pathway. Since Model A effectively subsumes the parameters in Model D, the same parameters were used for Model D, where appropriate. The output of 1000 iterations of Model D is shown in **Figure 10.**

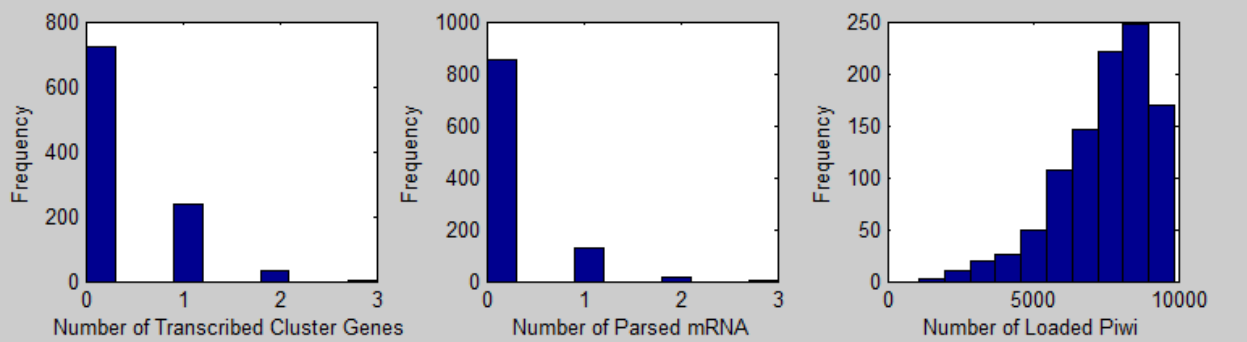

Figure 10: Outputted parameters of interest using the simulation for Model D. Of interest were the number of transcribed cluster genes, parsed mRNA generated from the cluster transcript, and the number of Piwi complexes loaded with parsed mRNA.

4 Discussion

4.1 Model A: Germline piRNA Pathway

Using the preliminary parameters in Model A, it was clear in most simulations there were low levels of transcribed genes and cleaved and parsed mRNA at the endpoint. Both Piwi and Aub were loaded at high frequencies in the majority of simulations and AGO3 had low numbers of loading. Because the Aub and Piwi were loaded highly but there were a much smaller number of Ping-Pong cycles undergone, this suggests that the majority of the loading was performed outside of the Ping-Pong cycle—in the initial steps. By this model, the Ping-Pong cycle would only perform a small amount of amplification of the initial impetus to silence active TE elements and thus not have as large of an impact as initially hypothesized.

Initial parameter modification allowed for new insights into the dynamics of the simulation as well as suggesting assumptions that may have been incorrect and areas in which the simulation could be improved. From this assessment, it was apparent that one of the key parameters was the rate of transcription of the *Max element* gene cluster. Looking at **Figure 4**, at the end of all simulations there were no cleaved mRNA available, which suggests that the pathway's progress was being limited by the lack of this necessary reactant for the progression of the Ping-Pong cycle.

4.2 Model B: Germline piRNA Pathway—Ping-Pong Intensive

The key differentiator between Model A and Model B was the transcription rate constant corresponding to the *Max element* gene. In simulations corresponding to the initial setup with this parameter it was evident that the number of cleaved *Max element* mRNA was a limiting factor in how many Ping-Pong cycles could occur. In Model A, all 1000 simulations had exactly 0 cleaved mRNA at the end-time point, which is why further study of this pathway necessitated parameter modification. In Model B, the rate constant was increased 1000-fold to alleviate this problem. The results from Model B shown in Figure 6 implicates that even with this dramatically increased rate constant, the number of cleaved mRNA was still 0, which further suggests that this cycle was taking in the cleaved mRNA faster than they could be regenerated. Of additional note is that while the number of loaded AGO3 remained at a similar range, the number of Ping-Pong cycles increased dramatically to an average nearly 100 times higher. This increased Ping-Pong activity in the same amount of simulation time provides a compelling reason why the mean number of loaded Piwi and Aub (which are loaded as a result of the Ping-Pong cycle) increased even further to the point where all 1000 simulations had all 10,000 Piwi and Aub loaded, with 0 variance.

Overall, from comparison of Model A and Model B we see that while given the initial simulation parameters, the number of Piwi—and thus the number of targeted TEs—was initially high, increasing the number of Ping-Pong iterations possible by adjusting the unknown parameters made this value even higher. While doing so set this parameter outside of the likely range for transcription initiation (the limiting factor in transcription), a similar effect would be noticed if there were more than one gene or higher polymerase activity associated with this gene.

4.3 Model C: Germline piRNA Pathway—Ping-Pong Intensive, AGO3 Competition

In the paper by Senti et al., it was suggested that AGO3 was potentially loaded with mRNA transcripts that were parsed from the initial *cluster* gene; however, these did not necessarily follow the same trends or mechanisms displayed by the initial loading of Piwi or Aub. Additionally, the function of these loaded AGO3 complexes was unclear as the Senti group did not suggest a subsequent use for them later in the pathway or in related pathways. Here, I suggest that despite not having a clear guess at the function of these loaded AGO3 complexes, one potential implication is that they would competitively inhibit the loading of AGO3 with cleaved *max element* mRNAs within the Ping-Pong cycle. Thus, I designed and tested a new model, Model C, in which AGO3 could face competitive inhibition by these parsed *cluster* genes. To do so, I modified Model A and created a new parameter that counted the number of these "blocked" AGO3 complexes. This simulation was run under 3 conditions corresponding to 3 variations of the rate parameter of *cluster* mRNAs binding to AGO3 to yield blocked AGO3. The initial setup detailed in **Figure 7** corresponded to the case where this parameter was equivalent to that of the parsed *max element* mRNAs being loaded onto AGO3. That of **Figure 8** and **Figure 9** correspond to the cases of a 1000 fold increase and decrease respectively. The equivalent case was largely similar to that of Model A, and there was minimal effect on the parameters. Intuitively, in this case, there would be comparable binding of AGO3 to either of the mRNA kinetically, and since the "loading" reaction occurs downstream of where the "blocking" reaction would occur, there is a much higher proportion of blocked mRNA compared to loaded. However, given an excess of AGO3, this does not appear to substantially affect the simulation. However in the case of the 1000 fold increase of the binding parameter, many of the parameters were affected. The number of Ping-Pong cycles dropped dramatically, and the number of loaded Piwi decreased to a fraction of what it was in the prior simulation. This suggests that while if there is comparable binding affinity to the two molecules it will not have a substantial effect, if there is preferential binding to the *cluster* mRNA, it would be expected that the Ping-Pong cycle would be adversely affected. Additionally, as one may expect, when the rate constant was decreased by 1000 fold, the simulation became, for the post part, roughly identical to that of Model A due to the very low probability of yielding a blocked AGO3 complex.

4.4 Model D: Somatic Pathway (No Ping-Pong)

While the previous models and modifications have demonstrated some of the intricacies and potential implications of the complexes involved in the Ping-Pong pathway, the creation of Model D—the "no Ping Pong" model—allows for a comparison to the effectively null model. In the somatic pathway, where the Ping-Pong cycle is not present, there are key differences as described earlier. For the sake of generation of Model D; however, these differences were simplified to the absence of the Aub and AGO3 complexes in favor of a linear model of Piwi loading for subsequent gene targeting. Instead of looking at 8 or 9 parameters of interest like the previous models, there were only 3 in this one: the number of transcribed cluster genes, the number of parsed mRNA, and the number of loaded Piwi. If the hypothesis that the Ping-Pong cycle acts as a positive feedback loop to target active TEs were true, one would expect that the number of loaded Piwi would, on average, be less than that of Model A, where similar parameters were used. Indeed, a comparison of **Figure 4** and **Figure 10** makes this point apparent—though not quite to the extent that one may anticipate when picturing a positive feedback or amplification loop. While this rough simulation does not provide proof that the Ping-Pong pathway acts as a positive feedback loop, it suggests

that—according to these parameters and simplifications—the Ping-Pong loop would be expected to increase the number of loaded Piwi complexes relative to the null control. Additionally, assuming that these loaded Piwi complexes solely serve the function of targeting corresponding TE sequences and that the rate of TE sequence targeting is proportional to the number of Piwi complexes for a simplified closed cell with confined spatial constraints, it would be logical to suggest that a function of the Ping-Pong pathway would be to increase the relative proportion of loaded Piwi complexes that target corresponding active TE sequences.

4.5 Sources of Error

In all of these simulations, one of the most concerning elements is the lack of actually measured parameters in the native system *Drosophila*. The lack of these elements prevents many of these conclusions from being transferrable to the native system. Once these parameters are experimentally measured, much more thorough and predictively valid models regarding the Ping-Pong system may be constructed. Additionally, the kinetics of many of these reactions outside of the transcription-based ones were not well-characterized in the literature, and thus many educated guesses and assumptions had to be made in the coding of these reactions for use in the Gillespie procedure. While simple MATLAB scripts such as the one used in the generation of these 4 models are useful for preliminary analyses and understanding the dynamics of a Ping-Pong-like model, it lacks the predictive validity of many of the more advanced models using exact reaction parameters and whole-cell simulations.

5 Conclusions

PiRNAs form RNA-protein complexes and have been identified in varied organisms with different biogenesis pathways and modes of action. They are unique in that they are thought to be produced from long polycistronic RNA transcripts by a Dicer-independent mechanism in both mammals and *Drosophila.* However, even with recent insights, there is a degree of uncertainty regarding their biogenesis and molecular function of piRNAs. In the Ping-Pong model of piRNA biogenesis, the piRNA generation process is coupled with RNA interference, which is mediated by three Piwi proteins known as PIWI, AUB, and AGO3. In the Ping-Pong mechanism, primary piRNAs cleave sense transposon transcripts and simultaneously produce secondary piRNAs from the sense transposons that then cleave antisense transposon transcripts. This mechanism can be viewed as dependent on the transcription of both sense and anti-sense transposon transcripts. The Ping-Pong mechanism is a form of a positive feedback loop, though it is unknown how the mechanism is started in the first place. The available data on primary piRNA production is very limited and the proposed models even in the most recent literature are, as a result, highly speculative. In germline cells, Ping-Pong amplification is used in addition to the primary piRNA pathway. Clearly there are differences between the piRNA biogenesis in somatic and germline cells, but the absence or presence of the Ping-Pong pathway has not been fully explained. Previous works have provided experimental evidence of the importance of the amplification loop in defense against transposons as underlined by studies of hybrid dysgenesis in *Drosophila.* In this work, preliminary modeling was done using a simplified model of the Piwi biogenesis and Ping-Pong pathways. A total of four models were created to preliminarily understand the biogenesis of Piwi-interacting RNAs from a stochastic simulation standpoint. Two of these models focused on a reaction scheme that involved Ping-Pong and attempted to characterize its behavior. One model suggested implications of a system in which one of the key complexes, AGO3, underwent competitive inhibition. The fourth model characterized a system that did not have the Ping-Pong cycle and allowed for comparisons between the two biogenesis pathways and suggested some potential implications regarding the differences in function of the Piwi-related pathways in somatic and germline cells. While these models presented a preliminary understanding of the basics of a positive feedback or amplification loop roughly based on the dynamics of the Piwi system, there is a great deal of work to be done to gain a greater understanding of this system and a closer look at the dynamics of the Ping-Pong pathway.

References

[1] Yuh CG Lee and Charles H Langley, "Transposable Elements in Natural Populations of Drosophila Melanogaster," *The Royal Society*, vol. 365, pp. 1219-1228, 2010.

[2] Senthil M Kumar and Kevin C Chen, "Evolution of Animal Piwi-Interacting RNAs and Prokaryotic CRISPRs," *Briefings in Functional Genomics*, pp. 1-12, 2012.

[3] Akihiro Nagao et al., "Biogenesis Pathways of piRNAs Loaded Onto AGO3 in the Drosophila Testis," *RNA*, vol. 16, pp. 2503-2515, 2010.

[4] Jian Lu and Andrew G Clark, "Population Dynamics of PIWI-interacting RNAs (piRNAs) and Their Targets in Drosophila," *Genome Research*, vol. 20, pp. 212-227, 2010.

[5] Jaspreet S Khurana and William Theurkauf, "piRNAs, Transposon Siliencing, and Drosophila Germline Development," *JCB: Review*, pp. 905-913, 2010.

[6] Robert J Ross, Molly M Weiner, and Haifan Lin, "PIWI Proteins and PIWI-Interacting RNAs in the Soma," *Nature*, vol. 505, pp. 353-359, 2014.

[7] Kirsten-Andre Senti and Julius Brennecke, "The piRNA Pathway: A Fly's Perspective on the Guardian of the Genome," *Cell*, vol. 26, no. 12, pp. 499-508, 2010.

[8] Elie S Dolgin and Brian Charlesworth, "The Effects of Recombination Rate on the Distribution and Abundance of Transposable Elements," *The Genetics Society of America*, vol. 178, pp. 2169-2177, 2008.

[9] Daniel Gillespie, "Exact Stochastic Simulation of Coupled Chemical Reactions," *The Journal of Physical Chemistry*, vol. 81, no. 25, pp. 2340-2361, 1977.

[10] Mukund Thattai and Alexander van Oudenaarden, "Intrinsic Noice in Gene Regulatory Networks," *PNAS*, pp. 8614-8619, 2001.