Can we always sweep the details of RNA-processing under the carpet?

Filippos D Klironomos1, Juliette de Meaux2 and Johannes Berg1,3

1 Institute for Theoretical Physics, University of Cologne, Cologne, Germany
2 Institute for Evolution and Biodiversity, University of Münster; Münster, Germany
3 Systems Biology of Ageing—Sybacol, Cologne, Germany

E-mail: fklirono@uni-koeln.de

Received 22 April 2013
Accepted for publication 11 September 2013
Published 3 October 2013
Online at stacks.iop.org/PhysBio/10/056007

Abstract

RNA molecules follow a succession of enzyme-mediated processing steps from transcription to maturation. The participating enzymes, for example the spliceosome for mRNAs and Drosha and Dicer for microRNAs, are also produced in the cell and their copy-numbers fluctuate over time. Enzyme copy-number changes affect the processing rate of the substrate molecules; high enzyme numbers increase the processing rate, while low enzyme numbers decrease it. We study different RNA-processing cascades where enzyme copy-numbers are either fixed or fluctuate. We find that for the fixed enzyme copy-numbers, the substrates at steady-state are Poisson-distributed, and the whole RNA cascade dynamics can be understood as a single birth–death process of the mature RNA product. In this case, solely fluctuations in the timing of RNA processing lead to variation in the number of RNA molecules. However, we show analytically and numerically that when enzyme copy-numbers fluctuate, the strength of RNA fluctuations increases linearly with the RNA transcription rate. This linear effect becomes stronger as the speed of enzyme dynamics decreases relative to the speed of RNA dynamics. Interestingly, we find that under certain conditions, the RNA cascade can reduce the strength of fluctuations in the expression level of the mature RNA product. Finally, by investigating the effects of processing polymorphisms, we show that it is possible for the effects of transcriptional polymorphisms to be enhanced, reduced or even reversed. Our results provide a framework to understand the dynamics of RNA processing.

1. Introduction

The copy-number of enzymes that mediate particular reactions is a source of intrinsic fluctuations in gene expression [1, 2]. An enzyme chemically converts its substrate at a rate which may be either fixed or may change over time. However, even for a fixed rate of chemical conversions per enzyme, prolonged changes in the enzyme copy-number become noticeable at the level of the enzyme’s substrate. Consequently, the substrates convert at a rate that fluctuates over time. In this work, we consider the effect of such fluctuations on the production and processing of mRNA, microRNA (miRNA) or small interfering RNA (siRNA).

The principal enzyme involved in mRNA processing is the spliceosome, which removes intronic elements from precursor mRNA molecules [3, 4]. In the biogenesis of eukaryotic small RNAs (sRNA), encompassing miRNAs and siRNAs, one additionally finds (i) the microprocessor unit responsible for processing primary miRNAs, (ii) the RNA-dependent RNA (RDR) polymerase responsible for synthesizing complementary strands to single-stranded RNA (ssRNA) and (iii) the nuclease Dicer responsible for processing precursor sRNA molecules [5]. In the biogenesis of prokaryotic sRNA, one finds the cascade/cas enzymes performing similar processing steps [6]. The processed eukaryotic or prokaryotic sRNA molecules are then loaded to Argonaute proteins forming the so-called RNA-induced silencing complexes (RISCs), the units responsible for the post-transcriptional regulation (PTR) of mRNA transcripts [7, 8].

The enzymes involved in the mRNA and sRNA biogeneses are vital for the normal functioning of the cell.
and are often regulated by complex feedback networks. For example, miR162 targets the Dicer DCL1 mRNA in Arabidopsis thaliana, but the precursor of miR162 needs Dicer to mature [9]. Hence large expression levels of Dicer lead to large levels of the miRNA that in turn down regulates Dicer. However, the presence of the tight regulation of the enzymatic expression levels does not completely eliminate fluctuations in their copy-numbers. At best, regulation provides a mechanism to limit the range of fluctuations around the mean value of enzymatic copy-numbers at steady-state.

Studies of the protein biogenesis usually assume that mRNA is produced via a birth–death process [10, 11]. Studies of PTR involving sRNAs assume the same [12–16]. This assumption that the multi-step process of RNA biogenesis can be modeled by a single birth–death process of the mature RNA product requires examination. We address this question by studying the mRNA and sRNA biogeneses using models that explicitly include the sequential processing of RNA precursors by the different enzymes. We find that under certain conditions one can indeed replace the complex RNA-processing cascade by a single production process of constant rate. In that case, mature RNA production events are statistically independent and follow a homogeneous Poisson process. However, when these conditions are not met, this simple picture breaks down, and the range of fluctuations introduced statistical dependences in the RNA-processing events.

2. Results

We start with the simplest scenario possible, a chain of RNA-processing steps. Such RNA-processing cascades arise in different contexts.

2.1. mRNA biogenesis

The biogenesis of messenger RNA starts with the transcription from the DNA of the primary RNA transcript (pre-mRNA) and continues as the spliceosome excises introns until the final mRNA product is reached. If we assume that each of these steps takes place at a constant rate, the cascade of events can be depicted schematically as follows:

\[
\begin{align*}
\emptyset & \xrightarrow{k_0} \text{pre-mRNA} \xrightarrow{d_1} \emptyset, \\
\text{pre-mRNA}_n & \xrightarrow{k_{n-1}} \text{pre-mRNA}_{n-1} \xrightarrow{d_{n-1}} \emptyset, \\
\text{pre-mRNA}_1 & \xrightarrow{k_0} \text{mRNA} \xrightarrow{d_0} \emptyset,
\end{align*}
\]

where \(k_0\) is the gene transcription rate, \(k_{n-1}, \ldots, k_0\) are the precursor mRNA-processing rates and \(d_n, \ldots, d_0\) are the basal degradation rates of all mRNA products. The length of the cascade is determined by the number of introns \(n \) of the primary transcript. We will show that at steady-state and under certain conditions, each of the components of the mRNA cascade is Poisson-distributed, and the mRNA creation dynamics follows a homogeneous Poisson process. In that case, (1) can be replaced by \(\emptyset \xrightarrow{k_0} \text{mRNA} \xrightarrow{d_0} \emptyset\) if \(k_0\) is chosen to match the effective production rate of mRNA at steady-state in (1).

2.2. siRNA biogenesis

A second example of a chain of RNA processing is the biogenesis of endogenous siRNA. This process starts with the transcription of genes or transposable elements to ssRNA. The ssRNA is then converted by the RDR to double-stranded RNA (dsRNA), which in turn is cleaved by Dicer nuclease to mature siRNAs. The biogenesis of exogenous siRNAs, involved for example in infections or transfections, is delocalized with only the last processing step occurring at the place where the siRNAs operate. In any case, if we assume that each step takes place at a constant rate, then the siRNA cascade can be depicted as follows:

\[
\begin{align*}
\emptyset & \xrightarrow{k_2} \text{ssRNA} \xrightarrow{d_2} \emptyset, \\
\text{ssRNA} & \xrightarrow{k_1} \text{dsRNA} \xrightarrow{d_1} \emptyset, \\
\text{dsRNA} & \xrightarrow{k_0} \text{siRNA} \xrightarrow{d_0} \emptyset.
\end{align*}
\]

This reaction scheme is analogous to the mRNA-processing chain (1) and can be simplified under certain conditions to a single birth–death process as discussed below.

The biogenesis of trans-acting small interfering RNA (tasiRNA) is not included in this scheme. For this particular class of siRNAs, the assumption of the constant rate of production \(k_2\) of ssRNA in (2) is invalid: the tasiRNA biogenesis initiates when fragments of post-transcriptionally regulated mRNAs, cleaved with the help of miRNAs, are converted to dsRNA by the RDR [17, 18]. This process can fluctuate considerably, resulting in temporal changes in the production rate \(k_2\) of ssRNA [16].

2.3. miRNA biogenesis

miRNAs originate from either intergenic DNA regions having their own promoters or intragenic DNA regions of introns of protein-encoding genes [5]. The latter can vary in length: short intragenic fragments called mirtrons are processed directly by the spliceosome [19], whereas longer intragenic fragments are additionally processed by the microprocessor unit in animals [20] and by Dicer in plants [17, 20]. There is evidence that the processing order might matter, and also that the microprocessor unit and the spliceosome collaborate in their tasks [21]. The latter suggests that these two processing steps might not be entirely independent. When the microprocessor unit (or Dicer in plants) acts first, the intragenic miRNA biogenesis can be depicted as follows:

\[
\begin{align*}
\emptyset & \xrightarrow{k_2} \text{pri–miRNA}_{\text{pre–mRNA}} \xrightarrow{d_2} \emptyset, \\
\text{pri–miRNA}_{\text{pre–mRNA}} & \xrightarrow{k_1} \text{pre–mRNA} \xrightarrow{d_1} \emptyset, \\
\text{pre–mRNA} & \xrightarrow{k_0} \text{mRNA} \xrightarrow{d_0} \emptyset,
\end{align*}
\]

where \(\text{pri–miRNA}_{\text{pre–mRNA}}\) denotes the long RNA hairpin consisting of the primary miRNA intragenic element and the precursor mRNA fragment. If the spliceosome first acts on the long RNA hairpin, it produces pri-miRNA from the intronic fragment, which is consequently processed into pre-miRNA by the microprocessor unit (or Dicer in plants). Mirtrons have a biogenesis identical to (3), but have shorter pri-miRNAs and the processing of the long primary RNA hairpin is performed by the spliceosome only [19].
Intergenic miRNA follows a similar biogenesis as intragenic miRNA:

$$\varnothing \xrightarrow{k_z} \text{miRNA} \xrightarrow{d_z} \varnothing.$$ (5)

The only difference between intergenic and intragenic miRNA biogenesis is the additional production of miRNA for the latter after the spliceosome excises all introns from the pri-miRNA. The processing steps according to intermediate components. However, 

$$\text{product. We choose this representation because it separates the precursor and the other step involving the creation of the new miRNA hairpin. Both (3) and (4) have a structure similar to the mRNA cascade (1).}$$

Here, the processing steps are broken down into two sub-steps: one step involving the destruction of the ancestor precursor and the other step involving the creation of the new product. We choose this representation because it separates processes according to intermediate components. However, both processing reactions in (5) are not independent but rather take place simultaneously. For example, the process $x \xrightarrow{k_x} \varnothing$ never takes place without the partner reaction $x \xrightarrow{k_y} y$.

Below we show analytically that at steady-state, the $x$, $y$ and $z$ products of (5) are Poisson-distributed. The rest of the details in (1)–(4) not included in (5), for example the number of introns excised in (1), influence the average steady-state expression level of the mature product, i.e., the mean value of the Poisson distribution. All, else being equal, the expression levels of two mature miRNAs, for example, might be different if one has introns in its precursor and the other does not, but they are both going to be Poisson-distributed at steady-state if processing rates per precursor mRNA are constant.

2.5. Solution of the RNA biogenesis under constant reaction rates

We consider the simplest scenario in some detail here as the same tools will be used to treat the effects of fluctuating reaction rates in section 2.7. The master equation describing the generic RNA biogenesis (5) is

$$\frac{d\rho_x \rho_y \rho_z}{dt} = [k_x (E^-_x - 1) + d_x (E^+_x - 1)]x + k_z (E^+_x E^-_y - 1)y + d_z (E^-_z - 1)z \rho_{x,y,z},$$

where $\rho_{x,y,z}(t)$ is the probability for the molecular numbers in the system at a given time to be $x$, $y$, $z$. The shift operators $E^\pm$ are defined by $E^\pm g(n) = g(n \pm 1)$, where $n \in \{x, y, z\}$. Using the generating function $f(r, s, q, t) = \sum_{x,y,z} r^x s^y q^z \rho_{x,y,z}(t)$ in (6), we arrive at

$$\frac{\partial f(r, s, q, t)}{\partial t} = k_x (r - 1) f(r, s, q, t) + \frac{\partial f(r, s, q, t)}{\partial r}$$

$$+ [d_x (1 - r) + k_x (s - r)] \frac{\partial f(r, s, q, t)}{\partial r}$$

$$+ [d_y (1 - s) + k_y (q - s)] \frac{\partial f(r, s, q, t)}{\partial s}$$

$$+ d_z (1 - q) \frac{\partial f(r, s, q, t)}{\partial q}.$$ (7)

At steady-state, the product of three generating functions of Poisson distributions $f(r, s, q) = \mathcal{E}^{(r-1)} \mathcal{E}^{(s-1)} \mathcal{E}^{(q-1)}$ solves (7) when

$$\langle x \rangle = \frac{k_x}{d_x + k_y}, \quad \langle y \rangle = \frac{k_y}{d_y + k_z}, \quad \langle z \rangle = \frac{k_z}{d_z}.$$ (8)

Consequently, the steady-state distribution of $z$ in the cascade (5) is identical to the steady-state distribution of $z$ in the single birth–death process $\varnothing \xrightarrow{k_z} z \xrightarrow{d_z} \varnothing$, where $\tilde{z}_z = k_z$. As expected for Poisson-distributed quantities, we find $\langle n^2 \rangle - \langle n \rangle^2 = \langle n \rangle$ for $n = \{x, y, z\}$, and computing the Fano factor $F_n$, defined as the variance over the mean of the random variable $n$, we obtain $F_x = F_y = F_z = 1$.

So far, we have proved that at steady-state the mature product of (5) follows Poisson statistics. However, we have not shown that the production statistics of $z$ at steady-state in (5) is identical to the production statistics of $z$ for a constant effective rate $\tilde{k}_z$. This question will be addressed in section 2.8 after we investigate the case of RNA biogenesis with fluctuating processing rates.

2.6. Synergistic or antagonistic effects of polymorphisms in sRNA biogenesis

Variation across organisms in sRNA-processing dynamics can be twofold: (i) the same precursor sRNA can be processed
differently by the RNA-processing enzymes producing several isoforms of the mature sRNA [5] or (ii) different precursor sRNAs from different loci or different alleles across species can have different processing rates but produce identical mature sRNAs [25–28]. In the first case, the processing variation affects the efficiency of the recruitment of the mature sRNA isoforms by the Argonaute proteins [29], and consequently the recycling rate of recruited mature sRNAs after they have catalyzed a transcript-targeting event [7, 30]. Both of these effects have been addressed elsewhere [16]. Here, we investigate the case of differences in the processing rates between two precursor sRNAs that give rise to the identical mature sRNA products.

We assume that two miRNA alleles (miR\textsubscript{1,2}) produce the identical mature miRNA products, but have differences in the transcription and processing rates of their respective pri-miRNAs. In particular, we assume that miR\textsubscript{1} is transcribed at a faster rate than miR\textsubscript{2} (k\textsubscript{1} > k\textsubscript{2}), but the primary transcript of miR\textsubscript{1} (pri-miR\textsubscript{1}) is processed at a slower rate (k\textsubscript{1} < k\textsubscript{2}). Is it possible for the steady-state copy-number z\textsubscript{1} of the mature miRNA product of miR\textsubscript{1} to be less than z\textsubscript{2}, the identical corresponding mature miRNA product of miR\textsubscript{2}? The ratio of the steady-state expression levels of (8) yields

$$ \frac{z_1}{z_2} = \frac{k_1}{k_2} \frac{1 + d_x/k_{x_2}}{1 + d_x/k_{x_1}}. $$

As long as pri-miR\textsubscript{1} is processed at a lower rate than pri-miR\textsubscript{2} (k\textsubscript{1} < k\textsubscript{2}), and despite the fact that miR\textsubscript{1} is transcribed at a higher rate than miR\textsubscript{2} (k\textsubscript{1} > k\textsubscript{2}), the steady-state expression level of mat-miR\textsubscript{1} can still be higher (z\textsubscript{1} < z\textsubscript{2}). Similar results are obtained if variation is present in the processing rates of the precursor miRNAs instead of the primary miRNAs. In other words, the effects of polymorphisms in the transcription of sRNAs can be reversed or enhanced with the appearance of polymorphisms in their processing steps.

2.7. RNA biogenesis with fluctuating processing rates

So far, we considered mRNA and sRNA biogeneses with constant reaction rates and showed that at steady-state the mature RNA follows Poisson statistics. Now we consider changes in the processing rates due to fluctuations in enzyme copy-numbers. We work with the generic RNA cascade (5).

We define at a given instant the number of molecules of the enzymes that process RNA transcripts at the first and second steps in (5) as α and β, respectively. By taking into account variations in enzyme copy-numbers, the RNA biogenesis becomes

$$ \begin{align*}
\varnothing & \overset{k_\alpha}{\longrightarrow} x & \overset{k_{\alpha x}}{\longrightarrow} & \varnothing, \\
x & \overset{k_{\alpha x}}{\longrightarrow} y & \overset{k_{\beta y}}{\longrightarrow} & \varnothing, \\
y & \overset{k_{\beta y}}{\longrightarrow} z & \overset{d_z}{\longrightarrow} & \varnothing,
\end{align*} $$

(10)

where the replacements k\textsubscript{r} → k\textsubscript{r}α and k\textsubscript{r} → k\textsubscript{r}β are made in (5). That is, k\textsubscript{r} is now the constant conversion rate per x molecule and per α enzyme. However, the processing rate k\textsubscript{r}α(t) per x molecule fluctuates now over time. The same applies for the processing rate k\textsubscript{r}β(t) per y molecule.

We denote as 1/k\textsubscript{r} the characteristic timescale over which the enzyme expression level varies. If this characteristic timescale is much slower than the dynamics associated with the biogenesis of RNA, then we can use the results of (8) for the given values of α, β and ensemble-average over the equilibrium distributions of α, β. The Fano factors F\textsubscript{n} for n = \{x, y, z\} are given by

$$ F_n = \frac{\mathbb{E}_{\alpha,\beta}[\langle n^2 \rangle] - \mathbb{E}_{\alpha,\beta}[\langle n \rangle]^2}{\mathbb{E}_{\alpha,\beta}[\langle n \rangle]} = 1 + \frac{k_{\alpha}}{d_y} \mathbb{F}[S_n], $$

(11)

where \mathbb{E}_{\alpha,\beta}[\cdot] indicates averaging over the equilibrium distributions of α, β, and \mathbb{F}[S_n] are the Fano factors of the stochastic variables S\textsubscript{n} listed below:

$$ S_x = \frac{1}{1 + k_{\alpha}d_y}, $$

$$ S_y = k_{\alpha}d_x \frac{1}{1 + k_{\alpha}d_x + k_{\beta}d_y}, $$

$$ S_z = k_{\beta}d_z + 1 + k_{\alpha}d_x + k_{\beta}d_y. $$

(12)

No assumption has been made about the particular form of the distributions of α, β. The only assumptions are (i) the distribution of enzyme expression levels is in steady-state and (ii) the enzyme biogenesis dynamics is slower than the RNA biogenesis dynamics, allowing us to employ the adiabatic approximation. Formally, the latter condition is expressed as θ ≪ max\{k_{\alpha}κ, k_{\beta}κ, d_x, d_y, d_z\}.

How do fluctuations in α, β introduce a linear dependence on k\textsubscript{r} in (11)? The answer lies in the way enzyme copy-number fluctuations imitate on substrates. Any change in α or β during time intervals that are similar or longer than the typical time interval of substrate biogenesis is felt by the corresponding substrate. The strength of this change is always related to the abundance of the corresponding substrate which in turn is proportional to k\textsubscript{r}. Therefore, the more abundant the substrate, the larger the effect of enzyme copy-number fluctuations on them as well. Furthermore, the strength of this effect depends also on the timescale of enzyme biogenesis. For example, faster enzyme copy-number fluctuations have less effect over the longer timescales of substrate biogenesis.

In this case, the processing of the substrate seems to be taking place under almost constant rates and the Fano factors of (11) tend to unity. On the other hand, slower fluctuations in the enzyme biogeneses induce slower changes in the processing rates, which are perceived by the substrates and render their biogeneses more noisier.

The theoretical results of (11)–(12) for the strength of fluctuations in the RNA cascade are independent of the details of the enzyme biogenesis. However, for simplicity, we limit the enzyme biogenesis in numerical simulations to the following birth–death processes:

$$ \begin{align*}
\varnothing & \overset{A}{\longrightarrow} x \overset{\alpha x}{\longrightarrow} \varnothing, \\
\varnothing & \overset{B}{\longrightarrow} \beta \overset{\beta x}{\longrightarrow} \varnothing,
\end{align*} $$

(13)

where A and B are the enzymatic expression levels at steady-state and 1/k\textsubscript{r} governs the timescale during which
Figure 1. Strength of fluctuations in RNA biogenesis depends linearly on the RNA transcription rate. The Fano factors of x, y and z in (10) are plotted as functions of the transcription rate $k_x$. The parameter values used are typical for the siRNA biogenesis in Salmonella [32]: $k_A = k_B = d_1 = d_2 = 1/h, A = B = 500$. In the main plot the rate $k$ of the biogeneses of the RNA-processing enzymes is varied. Blue lines correspond to $k = 0.1$ h$^{-1}$, brown lines correspond to $k = 1$ h$^{-1}$ and red lines to $k = 10$ h$^{-1}$. The horizontal black line indicates the range of unity Fano factors predicted by (6) for Poissonian statistics. Variation in enzyme copy-numbers induces fluctuations in the substrates, whose amplitude depends linearly on the RNA transcription rate. For $k \geq 0.1$ h$^{-1}$ and for $d_1 > 1$ min$^{-1}$ we find $F_x < F_y < F_z$: the RNA cascade buffers the fluctuations in $z$. However, in the inset, we plot substrate Fano factors for twice as stable mature RNA ($d_2 = 0.5$ h$^{-1}$) and for $k = 0.1$ h$^{-1}$ and find $F_x < F_z < F_y$: the RNA cascade amplifies the strength of fluctuations in $z$. In the other hand, the average processing rate of this stable mature RNA is also reduced ($k_B = 0.1$ h$^{-1}$), the relation $F_x < F_z < F_y$ is restored (red lines) and the RNA cascade buffers again the fluctuations in $z$.

Fluctuations in the intermediate product $y$ are expected to be stronger than the fluctuations in the transcript $x$, because both the birth and death rates of $y$ vary with time, whereas only the death rate of $x$ varies with time. However, the birth rate of the mature product $z$ closely follows the fluctuations that $y$ undergoes. One would expect the possibility of arising $F_y > F_z$ as well. These observations are reflected in (12), where $S_y$ depends only on $\alpha$, whereas both $S_x$ and $S_z$ depend on $\alpha$ and $\beta$. In fact, $S_x$ and $S_z$ differ only by the term $S_x/S_z = k_\beta/d_1$. This term, along with the speed of enzyme fluctuations $k_x$, determines the differences observed among $F_y$ and $F_z$ in figure 1. For a given $k_x$, if $k_\beta/d_1 \gg 1$, then fluctuations in the ancestor substrates are amplified in the mature RNA product, whereas noise filtering takes place if $k_\beta/d_1 \ll 1$. During a given time interval and when $k_\beta \gg d_1$, there are on average many more $z$-production than $z$-degradation events and noise amplification occurs. On the other hand, when $k_\beta \ll d_1$, the $z$-degradation events overwhelm the $z$-production events resulting into noise filtering. This effect is shown in the inset of figure 1, which shows the results for $k_x = 0.1$ h$^{-1}$ and a twice as stable as the mature RNA product ($d_2 = 0.5$ h$^{-1}$) compared to the main plot. In this case, we observe $F_y \simeq F_z < F_x$: the RNA cascade amplifies the strength of fluctuations in the mature RNA product (blue lines). However, when the processing rate is reduced $k_\beta \ll d_1$, the situation reverses again (red lines) and $F_y \simeq F_z > F_x$: the RNA cascade buffers the noise in the expression level of the mature RNA product as is also observed in the main plot of the figure.

The majority of the mature RNA products are expected to be more stable than their precursors. For example, mature miRNAs when loaded to Argonaute proteins are stabilized, and in certain cases in vitro half-lives become longer than a day [36, 37]. Mature transcripts are also protected by 5’-capping and polyadenylation. Thus for most cases we expect $d_1 < [d_x, d_y]$, leading to the amplification of fluctuations through the processing cascade. On the other hand, in the cases where the processing rate per precursor molecule is lower than $d_x$, or the mature RNA product is unusually unstable, the RNA cascade will operate in the reverse regime and reduce the strength of fluctuations in the expression level of the mature RNA product.

The key result of equation (11) is that when RNA-processing enzyme copy-numbers fluctuate slower than their substrates, a gene’s transcription rate affects the copy-number fluctuations of the mature product, whether it is protein or small RNA. As we are going to show, this remains true even when transcription takes place in bursts due to fluctuations in the promoter state [1, 2]. As the copy-numbers of processing enzymes affect the processing of the products of several genes, fluctuations of processing enzymes can be viewed as a contribution to ‘extrinsic stochasticity’ of the gene expression [1, 2]. Below, we discuss how to pick up this signature experimentally.

2.8. Dynamics of production events in the RNA cascade

We have shown that in the absence of enzymatic copy-number fluctuations any product of the cascade (5) is Poisson-distributed at steady-state. We now show that the dynamics
of z-production in (5), and thus correlations to any order, are identical to the dynamics and correlations of a homogeneous Poisson process of rate $\tilde{k}_z$. For simplicity, let us investigate a variation of (5) consisting only of a single processing step

$$\varnothing \xrightarrow{k_z} x \xrightarrow{k_y + d_y} \varnothing.$$  \hspace{1cm} (14)

The addition of more processing steps is straightforward to handle. Since enzyme copy-numbers in (14) are fixed, x and y are Poisson-distributed at steady-state. Additionally, the dynamics of $x$ is a birth–death process $\varnothing \xrightarrow{k_z} x \xrightarrow{k_y + d_y} \varnothing$ with constant birth and death rates. In other words, the creation of $x$ molecules is a homogeneous Poisson process. Once an $x$ particle is created, it either decays or after a time interval it is converted into a y molecule. For constant processing rates, this time interval is exponentially distributed and the time instants of y-creation events become uniformly distributed. Considering all such y-creation events originating from the rest of the $x$ molecules, introduces random shifts into the time intervals of y-creation events. However, the distribution of the time instants of y-creation events remains uniform, and the underlying y-creation process remains a homogeneous Poisson process with a rate of $k_y$. Therefore, the statistics of the process $\varnothing \xrightarrow{k_y} y$ becomes identical to the statistics generated by the process $\varnothing \xrightarrow{k_y} y$. The same logic applies if more steps are added in the processing cascade (14). As we discuss below, this implies that the autocorrelation function of any component of the cascade is identical to the autocorrelation function of a birth–death process at constant effective rates.

The situation is different when the copy-number of the processing enzyme $\alpha(t)$ fluctuates over time, leading to changes in the $x$-processing rate $k_z\alpha(t)$. All the $x$ molecules present at a given time are either processed at a higher or at a lower rate depending on the value of $\alpha(t)$. According to our discussion so far, within the time interval of the order of $1/\kappa$ where $\alpha(t)$ is constant, all y-creation events follow a homogeneous Poisson process of rate $k_y\alpha(t)$. As $\alpha(t)$ changes, however, this rate changes too, resulting in inhomogeneities in size of the order of $1/\kappa$ in the overall distribution of y-creation times. Consequently, the homogeneous Poisson process $\varnothing \xrightarrow{k_y\alpha(t)} y$ with uniformly distributed in time y-creation events produces different statistics than the process $\varnothing \xrightarrow{k_y\alpha(t)} y$.

2.9. Autocorrelation function of the mature RNA product

The autocorrelation function $C_z(\tau) = \langle z(t + \tau)z(t) \rangle - \langle z(t + \tau) \rangle \langle z(t) \rangle$ is a measure that identifies temporal correlations in the copy-number of the mature RNA product $z$ in the cascades (5) and (10). When $\tau \to 0$, one expects $z(t + \tau)$ to be highly correlated with $z(t)$. When $\tau \to \infty$, one expects of them to decorrelate, that is, one expects all ‘memory’ of the value of $z(t)$ to be lost when we resample $z$ at a much later time-point.

In figure 2, we plot $C_z(\tau)/C_z(0)$ for three different systems: (i) the simple multistep cascade (5) with constant reaction rates, (ii) a single birth–death process of constant rates $\varnothing \xrightarrow{\tilde{k}_z} z \xrightarrow{\tilde{d}_z} \varnothing$ with $\tilde{k}_z = k_z(y)$ and $\tilde{d}_z$ taken from (i) at steady-state and (iii) the full multistep cascade (10) with a speed of enzyme copy-number fluctuations determined by $\kappa = 0.1$ h$^{-1}$. For any range of $\tau$, there is no significant difference in the autocorrelation function between (5) and a single birth–death process of constant rates. In line with our previous discussion, differences appear when enzymatic fluctuations are included. In this case, enzyme fluctuations taking place over $1/\kappa$-sized time intervals affect the corresponding substrates. The memory of this effect across all $z$ is embedded in the autocorrelation function: temporal correlations persist over intervals of the order of $1/\kappa$.

2.10. Effects due to RNA-processing remain prominent in the presence of transcription bursts

Gene transcription is a process with a significant level of intrinsic noise already without fluctuations from RNA processing [38]. A strong promoter ensures that a gene is transcribed most of the time, and pauses occur over short intervals only. A weak promoter, on the other hand, results into longer pauses of transcription leading to strong fluctuations in the copy-number of primary transcripts. RNA processing takes place downstream of transcription. As a result, fluctuations in RNA processing adds to the fluctuations due to bursty transcription.
Figure 3. RNA processing with transcription bursts. Fano factors of the RNA substrates $x$, $y$, $z$ in the chain (10) are plotted as functions of the rate of transcription activation $k_x$. The rate of transcription inactivation is fixed at $k_{\text{off}} = 0.1 \text{ min}^{-1}$ [38]. When transcription is active the transcription rate is $k_t = 60 \text{ min}^{-1}$. The remainder of the parameters are identical to figure 1. In the absence of enzyme copy-number fluctuations ($\kappa = \infty$, dashed blue lines) and when $k_{\text{on}} \sim k_{\text{off}}$ transcription bursting leads to non-Poissonian statistics. This deviation from Poisson statistics is further enhanced by fluctuations in RNA-processing downstream of transcription ($\kappa = 0.1 \text{ min}^{-1}$, solid brown lines).

Figure 3 shows the Fano factors of the RNA substrates at a fixed transcription rate as a function of the rate of transcription activation in the absence (dashed blue lines) and in the presence (solid brown lines) of enzymatic fluctuations. As expected, Fano factors in the absence of enzymatic fluctuations collapse to unity in the limit of strong transcription activation [10], whereas in the presence of slow enzymatic fluctuations we recover the result of figure 1. Additionally, fluctuations in the number of processed RNA increase as the rate of transcription activation is reduced and transcription becomes more irregular. The signature of RNA processing remains prominent throughout the whole range of transcription activation values. This is despite the conservative assumption in our numerical simulations that enzyme copy-number fluctuations are Poissonian. If RNA-processing enzymes are produced in bursts as well [11], the effect on RNA processing would be stronger than what figures 1 and 3 show.

2.11. Fluctuations in RNA-processing impact on protein biogenesis

Protein production takes place in bursts due to multiple translation events of mRNA transcripts even when transcription is constitutive and the mRNA biogenesis is a birth–death process of constant rates [10]. Introducing RNA enzyme fluctuations in the mRNA biogenesis is expected to render protein production even more noisy. Here, we investigate within our conservative framework how much fluctuations in protein production increase due to fluctuations in the copy-numbers of the RNA-processing enzymes.

Figure 4. Effect of RNA processing on protein biogenesis. Fano factors of pre-mRNA ($x$), mRNA ($y$) and protein ($z$) are plotted as functions of the gene transcription rate $k_t$ in the absence ($\kappa = \infty$, dashed blue lines) and in the presence ($\kappa = 0.1 \text{ min}^{-1}$, solid brown lines) of spliceosome copy-number fluctuations. The translation rate per mRNA molecule is fixed at $k_v = 5 \text{ h}^{-1}$ and the rest of parameters are identical to figure 1. In the absence of spliceosome fluctuations the statistics of $x$, $y$ is Poissonian, but bursts of translation lead to non-Poissonian statistics of $z$ with a Fano factor $F_z$ independent of $k_t$ [10]. When spliceosome copy-number fluctuations are included, fluctuations in protein production increase linearly with $k_t$. Our results are conservative since fluctuations in ribosomal copy-numbers are ignored, and spliceosome fluctuations are taken to be Poissonian.

Figure 4 shows the Fano factors of pre-mRNA ($x$), mRNA ($y$) and protein ($z$) as functions of the gene transcription rate in the absence (dashed blue lines) and presence (solid brown lines) of spliceosome copy-number fluctuations. Fano factors of the RNA substrates collapse to unity as expected in the absence of enzymatic fluctuations, and the protein Fano factor becomes $1 + (k_{/d_z})/(1 + d_z/d_y) = 3.5$ and independent of the transcription rate [10]. In the presence of RNA-processing enzyme copy-number fluctuations, however, we find that the protein biogenesis is significantly affected and protein noise becomes linearly dependent now on the transcription rate. We stress again that these results are conservative because RNA-processing enzyme fluctuations are modeled as Poissonian. Additionally, copy-number fluctuations of the ribosomal units are not included in the simulation. Relaxing any of these two conditions would only amplify the strength of the effect shown in figure 4.

3. Conclusion

RNA biogenesis is a multi-step cascading process for protein-encoding transcripts and sRNAs alike. The output of one process becomes the input of the next one until the final mature product is reached. We showed that when reactions in the RNA cascades occur at constant rates, the mature products undergo single-step birth–death biogenesises of constant effective rates and are Poisson-distributed at steady-state. This simple picture breaks down when there are fluctuations in the copy-numbers of enzymes that mediate RNA processing.

\[ F_z = \frac{\sigma_z^2}{\langle z \rangle^2} \]
Enzymatic fluctuations induce fluctuations in the processing rates per corresponding substrate molecule. We showed that Fano factors of the RNA cascade’s products increase linearly with the transcription rate, irrespective of the form of the steady-state distribution of the copy-numbers of enzymes participating in the processing steps. In numerical simulations that include transcriptional and translational bursting, we find that this effect remains detectable in the presence of other sources of fluctuations, especially for the case of protein biogenesis.

PTR is a significant part of sRNA biogenesis. Mature sRNAs are recruited by the Argonaute proteins in order to form the RISCs [5]. RISCs are the units that not only mediate PTR, but also ensure the stability of sRNAs, rendering them important elements of the sRNA biogenesis. Additionally, early sRNA-processing steps can be affected in numerous ways via feedback regulation. One example is miR162 regulating DCL1; further examples include miR168 regulating the Argonaute protein AGO1 and miR403 regulating AGO2 [9]. Furthermore, there is abundant evidence of feedback regulation to sRNA genes by their transcription factor targets [39]. All of this dynamics remained outside the scope of our analysis, as it involves PTR rather than the sRNA biogenesis.

However, if feedback regulates the transcription of sRNA only, then the conclusions of our analysis are still applicable. If enzyme fluctuations are negligible, one can replace the sRNA cascade with a single birth–death process and incorporate the feedback regulation into the sRNA birth process. If, on the other hand, like in the case of miR162 and DCL1, feedback regulation takes place in sRNA processing, then clearly the processing details matter and need to be included. This can be the topic of a future study of feedback dynamics within the RNA-processing cascade. Finally, many RNA subclasses are not mentioned in this work. For example, piwi-interacting RNAs, small nucleolar RNAs or small nuclear RNAs; all have distinct biogeneses and functionalities [40]. If these RNA subclasses follow processing chains like (5) or the more general (10), then our results are also applicable for them.

Single-nucleotide polymorphisms affect the processing rate of precursors of sRNAs and impact on the expression level of the mature products [28]. Naturally, polymorphisms affecting transcription in sRNA also induce variation across species in the production of mature sRNAs. However, based on our analysis of the steady-state expression levels in the sRNA biogenesis, we predict that the effect of transcriptional variation can be enhanced, reduced or even reversed by the presence of variation in sRNA processing. This result shows that the functional consequences of modification in pre-miRNA expression and processing rates are not intuitive and should depend on the evolutionary forces acting on the final miRNA product. In situations where a miRNA is under positive selection to be expressed at higher steady-state levels, increased processing might provide a greater advantage than an increase in the rate of transcription, even if the mutation occurs in a genetic background with a weak transcription rate. Conversely, in the face of evolutionary constraints for maintaining steady-state levels, one might observe compensatory evolution between mutations altering processing and transcription rates. Within population, this might lead to stronger linkage disequilibrium between the upstream regulatory region and the region encoding the miRNA precursor.

To date, functional changes in miRNA encoding genes during evolution were seldom examined, but a few studies often showed that both miRNA transcription and processing rates vary [41, 42]. Evaluating the extent of compensatory evolution requires experimental systems where the degradation products of miRNA processing can be quantified so that transcription and processing rates can be estimated. Deep sequencing technologies are now making such estimates possible.

Our analysis showed that the RNA cascade amplifies the noise in the expression level of a mature RNA product, provided the mature product is more stable than its precursors (as is generally the case). However, we showed that it is also possible for the RNA cascade to reduce the noise in the expression level of unusually unstable mature RNA products or of those mature products whose precursors are processed at low rates.

The simplest experimental test of our results would involve a Tet-inducible promoter that drives the expression of a fluorescent protein, whose transcript contains at least one intron in order to ensure that the spliceosome will interact with it. Spliceosome copy-number fluctuations can be induced indirectly by specific inhibitors [43, 44]. The induced expression of the fluorescent protein under slow spliceosome copy-number fluctuations will result in much broader protein distributions due to the broadening of the mRNA distributions according to equation (11). Furthermore, we expect a stronger expression of the fluorescent protein under fixed spliceosome inhibitor concentrations to result also in broader protein distributions with Fano factors linearly depending on the tetracycline concentration as we showed in figure 4.

In summary, we investigated the RNA-processing cascade and found universal characteristics in the steady-state dynamics for different RNA species. If processing steps take place at constant rates, then the mature RNA biogenesis can be modeled at steady-state as a single birth–death process of constant rates. Variation in the processing rates induces additional fluctuations to the RNA cascade and the single birth–death picture breaks down. Finally, we showed that polymorphisms in the processing rates can act synergistically or antagonistically to polymorphisms in the transcription rates of RNA. Our work offers a framework to better understand the dynamics of RNA biogenesis and *a fortiori* of RNA evolution.

Acknowledgments

This work was supported by Deutsche Forschungsgemeinschaft (DFG) grant SFB 680, by the BMBF-SysMO2 grant and by SyBaCol.

References

[1] Swain P S, Elowitz M B and Siggia E D 2002 Intrinsic and extrinsic contributions to stochasticity in gene expression *Proc. Natl Acad. Sci. USA* 99 12795–800
[2] Paulsson J 2005 Models of stochastic gene expression Phys. Life Rev. 2 157–75
[3] Wahl M C, Will C L and Lührmann R 2009 The splicesome: design principles of a dynamic RNP machine Cell 136 701–18
[4] Will C L and Lührmann R 2011 Splicesome structure and function Cold Spring Harbor Perspect. Biol. 3 a003707
[5] Krolof J, Loedige I and Filippowicz W 2010 The widespread regulation of microRNA biogenesis, function and decay Nature Rev. Genet. 11 597–610
[6] van der Oost J and Stan J B 2009 RNAi: prokaryotes get in on the act Cell 139 863–5
[7] Jones-Rhoades M W, Bartel D P and Bartel B 2006 MicroRNAs and their regulatory roles in plants Annu. Rev. Plant Biol. 57 19–53
[8] Narry Kim V, Han J and Siomi M C 2009 Biogenesis of small RNAs in animals Nature Rev. Mol. Cell Biol. 10 126–39
[9] Allen E, Xie Z, Gustafson A M and Carrington J C 2005 microRNA-directed phasing during trans-acting siRNA biogenesis in plants Cell 121 207–21
[10] Thattai M and van Oudenaarden A 2001 Intrinsic noise in gene regulatory networks Proc. Natl Acad. Sci. USA 98 8614–9
[11] Shahrezaei V and Swain P S 2008 Analytical distributions for stochastic gene expression Proc. Natl Acad. Sci. 105 17256–61
[12] Shimoni Y, Friedlander G, Hetzroni G, Niv G, Altuvia S, Biham O and Margalit H 2007 Regulation of gene expression by small non-coding RNAs: a quantitative view Mol. Syst. Biol. 3 138
[13] Levine E and Hwa T 2008 Small RNAs establish gene expression thresholds Curr. Opin. Microbiol. 11 574–9
[14] Hao Y, Liufang Xu and Shi H 2011 Theoretical analysis of catalytic-sRNA-mediated gene silencing J. Mol. Biol. 406 195–204
[15] Wang S and Raghavachari S 2011 Quantifying negative feedback regulation by micro-RNAs Phys. Biol. 8 055002
[16] Klironomos F D and Berg J 2013 Quantitative analysis of competition in posttranscriptional regulation reveals a novel signature in target expression variation Biophys. J. 104 951–8
[17] Rajagopalan R, Vaucheret H, Trejo J and Bartel D P 2006 A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana Genes Dev. 20 3407–25
[18] Chen H-M, Chen L-T, Patel K, Li Y-H, Baulcombe D C and Wu S-H 2010 22-nucleotide RNAs trigger secondary siRNA biogenesis in plants Proc. Natl Acad. Sci. 107 597–610
[19] Westholm J O and Lai E C 2011 Mirtrons: microRNA biogenesis via splicing Biochimie 93 1897–902
[20] Zhu J-K 2008 Reconstituting plant miRNA biogenesis Proc. Natl Acad. Sci. 105 9851–2
[21] Janas M M, Khaled M, Schubert S, Bernstein J G, Golan D, Veguilla R A, Fisher D E, Shomron N, Levy C and Novina C D 2011 Feed-forward microprocessing and splicing activities at a MicroRNA-containing intron PLoS Genet. 7 e1002330
[22] Hinske L C G, Galante P A F, Kuo W P and Ohno-Machado L 2010 A potential role for intragenic miRNAs on their hosts’ interactome BMC Genomics 11 533
[23] Mukherji S, Ebert M S, Zheng G X Y, Tsang J S, Sharp P A and van Oudenaren A 2011 MicroRNAs can generate thresholds in target gene expression Nat. Genet. 43 854–9
[24] Gjuvsland A, Plath E and Omholt S 2007 Threshold-dominated regulation hides genetic variation in gene expression networks BMC Syst. Biol. 1 57
[25] Xie Z, Allen E, Fahlgren N, Calamar A, Givan S A and Carrington J C 2005 Expression of Arabidopsis miRNA genes Plant Physiol. 138 2145–54
[26] Li W, Cui X, Meng Z, Huang X, Xie Qi, Wu H, Jin H, Zhang D and Liang W 2012 Transcriptional regulation of Arabidopsis MIR168a and ARGONAUTE1 homeostasis in abscisic acid and abiotic stress responses Plant Physiol. 158 1279–92
[27] de Meaux J, Jin-Yong Hu, Tartler U and Goebel U 2008 Structurally different alleles of the ath-MIR824 microRNA precursor are maintained at high frequency in Arabidopsis thaliana Proc. Natl Acad. Sci. 105 8994–9
[28] Todesco M, Balasubramanian S, Cao J, Ott F, Suresh Kumar S, Schneebberger K, Meyer R C, Altmann T and Weigel D 2012 Natural variation in biogenesis efficiency of individual Arabidopsis thaliana MicroRNAs Curr. Biol. 22 166–70
[29] Juvvuna P K, Khandelia P, Lee L M and Makeeey E V 2012 Argonaute identity defines the length of mature mammalian microRNAs Nucleic Acids Res. 40 6808–20
[30] Haley B and Zamore P D 2004 Kinetic analysis of the RNAi enzyme complex Nat. Struct. Mol. Biol. 11 599–606
[31] Gillespie D T 1977 Exact stochastic simulation of coupled chemical reactions J. Phys. Chem. 81 2340–61
[32] Martin O, Jesper J, Jakob M J and Pouil V H 2009 Switching off small RNA regulation with trampRNA Mol. Microbiol. 73 790–800
[33] Schwannhüse L, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W and Selbach M 2011 Global quantification of mammalian gene expression control Nature 473 337–42
[34] Selbach M, Schwannhüse L, Thierfelder N, Fang Z, Khanin R and Rajewsky N 2008 Widespread changes in protein synthesis induced by microRNAs Nature 455 58–63
[35] Guo H, Ingolia N T, Weissman J S and Bartel D P 2010 Mammalian microRNAs predominantly act to decrease target mRNA levels Nature 466 835–40
[36] Baccarini A, Chauhan H, Gardner T J, Jayaprakash A D, Sachidanandam R and Brown B D 2011 Kinetic analysis reveals the fate of a MicroRNA following target regulation in mammalian cells Curr. Biol. 21 369–76
[37] Gantier M P et al 2011 Analysis of microRNA turnover in mammalian cells following Dicer1 ablation Nucleic Acids Res. 39 5692–703
[38] Goldberg I, Paulsson J, Zawilski S M and Cox E C 2005 Real-time kinetics of gene activity in individual bacteria Cell 123 1025–36
[39] Lutter D, Marr C, Krumsieck J, Lang E W and Theis F J 2010 Intrinsic microRNAs support their host genes by mediating synergistic and antagonistic regulatory effects BMC Genomics 11 224
[40] Guil S and Esteller M 2012 Cis-acting noncoding RNAs: friends and foes Nature Struct. Mol. Biol. 19 1068–75
[41] Danny W-K N, Zhang C, Miller M, Palmer G, Whiteley M, Tholl D and Chen Z J 2011 Cis- and trans-regulation of miR163 and target genes confers natural variation of secondary metabolites in two Arabidopsis species and their allotriploids Plant Cell 23 1729–40
[42] Wang Y, Ilaya A, Zhong X, Yang Wu, Zhang J, van der Knaap E, Olmstead R, Yijun Q and Ding B 2011 Function and evolution of a MicroRNA that regulates a Ca2+-ATPase and triggers the formation of phased small interfering RNAs in tomato reproductive growth Plant Cell 23 3185–203
[43] Kuhn J A, van Santen M A, Schwienhorst A, Urlaub H and Lührmann R 2009 Stalling of spliceosome assembly at distinct stages by small-molecule inhibitors of protein acetylation and deacetylation RNA 15 153–75
[44] Effgenberger K A, Perriman R J, Bray W M, Lokey R S, Ares M and Jurica M S 2013 A high-throughput splicing assay identifies new classes of inhibitors of human and yeast splicesomes J. Biomol. Screen. 18 1110–20