Derivation of the protein distribution for a two-stage model of gene expression

From the master equation

The generating function for the master equation of the two-stage model satisfies (Eq. 1 in the main text)

$$\frac{1}{v} \frac{\partial F}{\partial \tau} + \frac{\partial F}{\partial v} - \gamma \left[ b(1 + u) - \frac{u}{v} \right] \frac{\partial F}{\partial u} = a \frac{u}{v} F \quad (23)$$

where $F(z', z)$ is defined as $\sum_{m,n}(z')^m z^n P_{m,n}$, and we have let $u = z' - 1$ and $v = z - 1$. If $r$ measure the distance along a characteristic which starts at $\tau = 0$ with $u = u_0$ and $v = v_0$ for some constant $u_0$ and $v_0$, then Eq. 23 becomes

$$\frac{dv}{dr} = \frac{1}{v} ; \quad \frac{d\tau}{dr} = 1 ; \quad \frac{du}{dv} = -\gamma \left[ b(1 + u) - \frac{u}{v} \right] ; \quad \frac{dF}{dr} = \frac{au}{v} F. \quad (24)$$

Consequently, $v = r$ and

$$\frac{du}{dv} = -\gamma \left[ b(1 + u) - \frac{u}{v} \right] \quad (25)$$

which has solution

$$u(v) = e^{-\gamma bv} v^\gamma \left[ C - b\gamma \int_v^\infty dv' \frac{e^{\gamma bv'}}{(v')^{\gamma+1}} \right] \quad (26)$$

for a constant $C$ as can be verified by differentiation. By Taylor expanding $e^{\gamma bv}$ so that $e^{\gamma bv} = \sum_n \frac{(\gamma bv)^n}{n!}$, we can evaluate the integral in Eq. 26,

$$u(v) = e^{-\gamma bv} \left[ C v^\gamma - \sum_{n=0}^\infty \frac{(\gamma bv)^{n+1}}{n!(n - \gamma + 1)} \right]. \quad (27)$$

We can also carry out the sum in Eq. 27 in the limit of $\gamma \gg 1$ following Bender and Orzag [1]. By comparing the ratio of the $n - 1$’th and the $n$’th term, we see that the elements of the sum have a maximum when $n \simeq \gamma bv$. For $\gamma \gg 1$, the sum will be dominated by terms with $n$ near $\gamma bv$. We therefore let $n = \gamma bv + s$ for some $s$, then $n!$ can be shown to be approximately [1]

$$n! \simeq (\gamma bv)^n e^{-\gamma bv} e^{\frac{2s^2}{\gamma bv}} \sqrt{2\pi \gamma bv} \quad (28)$$
using Stirling’s approximation. Consequently, by approximating the sum as an integral and extending the range of the integral to $-\infty$, 

$$\sum_{n=0}^{\infty} \frac{(\gamma bv)^{n+1}}{n!(n - \gamma + 1)} \simeq \int_{-\infty}^{\infty} ds \frac{e^{-\frac{s^2}{2\pi \gamma bv}}}{\sqrt{2\pi \gamma bv}} \cdot \frac{\gamma bv e^{bv}}{\gamma (bv - 1) + s + 1}$$

$$= \int_{-\infty}^{\infty} ds \frac{e^{-\frac{s^2}{2\pi \gamma bv}}}{\sqrt{2\pi \gamma bv}} \cdot \frac{bv e^{bv}}{bv - 1} \left[ 1 + \gamma^{-1} \left( \frac{s + 1}{bv - 1} \right) \right]^{-1}$$

$$= \frac{bv e^{bv}}{bv - 1} \int_{-\infty}^{\infty} ds \frac{e^{-\frac{s^2}{2\pi \gamma bv}}}{\sqrt{2\pi \gamma bv}} + O(\gamma^{-1})$$

$$\simeq \frac{bv e^{bv}}{bv - 1}$$

(29)

to the lowest order in $\gamma$. From Eq. 27, $u$ satisfies

$$u(v) \simeq Ce^{-\gamma bv}v^\gamma + \frac{bv}{1 - bv}$$

(30)

when $\gamma \gg 1$. We evaluate $C$ using $u = u_0$ when $v = v_0$ giving

$$u \simeq \left( u_0 - \frac{bv}{1 - bv} \right) e^{-\gamma b(v - v_0)} \left( \frac{v}{v_0} \right)^\gamma + \frac{bv}{1 - bv}$$

$$\simeq \frac{bv}{1 - bv}.$$  

(31)

when $\gamma \gg 1$ because $v = v_0 e^\tau > v_0$ from Eq. 24.

**Finding the generating function**

Using Eq. 31, Eq. 24 becomes

$$\frac{dF}{dv} = \frac{ab}{1 - bv} F$$

(32)

or, on integrating,

$$\log F(v) F(v_0) = -a \log \left( \frac{1 - bv}{1 - bv_0} \right)$$

(33)

because $F(v_0) = F(\tau = 0)$. If initially we have $k$ proteins then

$$F(v_0) = \sum P_n(\tau = 0) z^n = \sum \delta_{n,k} z^n = z^k = (1 + v_0)^k.$$  

(34)

For our approximation, Eq. 31, to be valid, enough time must have passed for mRNA levels to have reached steady-state. Strictly, this initial condition is only valid for non-zero $\tau$ of the order of $d_1/d_0 = \gamma^{-1}$. Finally, inserting Eq. 34 into Eq. 33 gives

$$F(z, \tau) = \left[ \frac{1 - b(z - 1)e^{-\tau}}{1 - bz + b} \right]^a \left[ 1 + (z - 1)e^{-\tau} \right]^k$$

(35)

because $v_0 = (z - 1)e^{-\tau}$. When $k = 0$, Eq. 35 becomes Eq. 7.
Deriving the probability distribution for proteins

We can find $P_n(\tau)$, the probability of having $n$ proteins at time $\tau$ given initially zero proteins, by differentiating Eq. 35 when $k = 0$. By definition, $P_n$ satisfies $P_n = \frac{1}{n!} \frac{\partial^n}{\partial z^n} F(z, \tau) \big|_{z=0}$. By writing

$$F(z, \tau) = \left( \frac{1 + be^{-\tau}}{1 + b} \right)^a \cdot \left[ 1 - \frac{b}{1 + e\tau} \right]^{\frac{n}{1 + b} - a} \cdot \left[ 1 - \frac{b}{e^{a+\tau} + b} \right]^{-\frac{n}{1 + b} - a},$$  \hspace{1cm} (36)

we can make use of the identities

$$\frac{\partial^n}{\partial z^n} [1 - qz]^{-a} \bigg|_{z=0} = \frac{\Gamma(a + n)}{\Gamma(a)} q^n$$  \hspace{1cm} (37)

and

$$\frac{\partial^n x(z)}{\partial z^n} y(z) = n! \sum_{k=0}^{n} \frac{\partial^{n-k}}{\partial z^{n-k}} x(z) \cdot \sum_{j=0}^{k} \frac{(-1)^j (k + 1) y(z)^{j-1}}{(j + 1)! (n - k)! (k - j)!} \frac{\partial^k}{\partial z^k} z^j$$  \hspace{1cm} (38)

which is given at Wolfram Research (functions.wolfram.com/GeneralIdentities/9).

Interpreting $x(z)$ as the numerator of the quotient in Eq. 36 and $y(z)$ as its denominator, we find

$$P_n(\tau) = \left( \frac{1 + be^{-\tau}}{1 + b} \right)^a \sum_{k=0}^{n} \frac{\Gamma(a + n - k)}{\Gamma(a)} \left( \frac{b}{1 + b} \right)^{n-k} \times \sum_{j=0}^{k} \frac{(-1)^j (k + 1) y(z)^{-j+1}}{(j + 1)! (n - k)! (k - j)!} \frac{\partial^k}{\partial z^k} y(z)^j$$  \hspace{1cm} (39)

where we can use

$$\sum_{j=1}^{k} \frac{(-1)^j \Gamma(a j + k)}{\Gamma(a j) (j + 1)! (k - j)!} = \frac{(-1)^k \Gamma(a + 1)}{\Gamma(a - k + 1) (k + 1)!}$$  \hspace{1cm} (40)

to simplify further. Eq. 40 can be verified by directly expanding the sum. Consequently,

$$P_n(\tau) = \left( \frac{b}{1 + b} \right)^n \frac{\Gamma(a + n)}{\Gamma(a)} \sum_{k=0}^{n} \frac{(-1)^k}{k!} \frac{\Gamma(a - k + n)}{\Gamma(n - k + 1) \Gamma(a - k + 1)} \left( \frac{1 + b e^{-\tau}}{e^{a+\tau} + b} \right)^k.$$  \hspace{1cm} (41)

The hypergeometric function $2F_1(a, b, c; z)$ obeys

$$2F_1(-n, b, c; z) = \sum_{k=0}^{n} \frac{(-1)^k \Gamma(n + 1)}{\Gamma(n - k + 1) \Gamma(c) k!} \frac{(b)_k z^k}{(c)_k k!}$$  \hspace{1cm} (42)

when $a$ is a negative integer and where $(b)_k$ and $(c)_k$ are Pochhammer symbols [2]. From their definition, $(a)_k = \Gamma(a + k)/\Gamma(a)$, the Pochhammer symbols satisfy

$$\Gamma(a + 1) = (-1)^k (-a)_k \Gamma(a - k + 1).$$  \hspace{1cm} (43)

Writing $\Gamma(a - k + n) = \Gamma(a + n - 1 - k + 1)$ and using Eq. 42 and Eq. 43, we find that

$$P_n(\tau) = \frac{1}{n!} \left( \frac{b}{1 + b} \right)^n \left( \frac{1 + be^{-\tau}}{1 + b} \right)^a \frac{\Gamma(a + n)}{\Gamma(a)} 2F_1 \left( -n, -a, 1 - a - n; \frac{1 + b}{e^{a+\tau} + b} \right).$$  \hspace{1cm} (44)

which is valid for $\gamma \gg 1$, $\tau > \gamma^{-1}$, and $a$ and $b$ finite.
Deriving the ‘propagator’ probability

By differentiating Eq. 35 for non-zero \( k \), we can express the ‘propagator’ probability, \( P_{n|k}(\tau) \), in terms of Eq. 44. From the definition of \( P_n(\tau) \), Eq. 35 can be written as

\[
F(z, \tau) = \left[ \sum_{n=0}^{\infty} P_n(\tau) z^n \right] \left[ 1 - e^{-\tau} + z e^{-\tau} \right]^k
\]  

(45)

or

\[
F(z, \tau) = \sum_{n=0}^{\infty} P_n(\tau) z^n \sum_{r=0}^{k} \binom{k}{r} (1 - e^{-\tau})^{k-r} (ze^{-\tau})^r
\]  

(46)

using the binomial theorem. From the coefficients of the powers of \( z \), we find

\[
P_{n|k}(\tau) = \sum_{r=0}^{k} \binom{k}{r} P_{n-r}(\tau) (1 - e^{-\tau})^{k-r} e^{-r\tau}
\]  

(47)

because \( F(z, \tau) = \sum_n P_{n|k}(\tau) z^n \) and remembering that \( P_n(\tau) = 0 \) if \( n < 0 \).

Finding the probability distribution for the first passage time

With \( P_n(\tau) \) and \( P_{n|k}(\tau) \), we can find the distribution for the first time the number of proteins reaches a threshold \( N \). We define this distribution to be \( f_N(\tau) \). It obeys a renewal equation [3]

\[
P_N(\tau) = \int_{0}^{\tau} d\tau' f_N(\tau') P_{N|N}(\tau - \tau').
\]  

(48)

The probability of having \( N \) proteins at time \( \tau \) is equal to the sum of the probability of first reaching \( N \) proteins at \( \tau' \) and then returning to \( N \) proteins at a time \( \tau - \tau' \) later for all times \( \tau' \) less than \( \tau \). We have assumed that the initial number of proteins is zero, but this assumption is not necessary.

Eq. 48 is a Volterra integral equation of the first kind and can be straightforwardly solved numerically [4]. If \( N > 0 \) then \( f_N(0) = 0 \) and \( P_{N|N}(0) = 1 \) by definition. Consequently, by discretizing and letting \( \tau_i = i\epsilon \) for integer \( i \) and small \( \epsilon \), we can write the integral in Eq. 48 as a trapezium rule:

\[
\int_{0}^{\tau_i} d\tau' f_N(\tau') P_{N|N}(\tau_i - \tau') \simeq \epsilon \left[ \frac{1}{2} f_N(\tau_i) + \sum_{j=1}^{i-1} P_{N|N}(\tau_i - \tau_j) f_N(\tau_j) \right].
\]  

(49)

Inserting Eq. 49 into Eq. 48 gives a series of equations for \( f_N(\tau_i) \) which we solve iteratively:

\[
f_N(\tau_1) = \frac{2P_N(\tau_1)}{\epsilon}
\]  

(50)

\[
f_N(\tau_i) = 2 \left[ \frac{P_N(\tau_i)}{\epsilon} - \sum_{j=1}^{i-1} P_{N|N}(\tau_i - \tau_j) f_N(\tau_j) \right].
\]  

(51)

We implement Eqs. 50 and 51 in Matlab (The Mathworks, Natick, Massachusetts). Our code is available at www.cnd.mcgill.ca/~swain.

We use

\[
\langle n(\tau_1) n(\tau_2) \rangle = \sum_{n,n'} n n' P_{n|n'}(\tau_2 - \tau_1) P_{n'}(\tau_1)
\]  

(52)

to find the auto-correlation function. We evaluate the sum in Eq. 52 numerically, cutting off the sums when \( n \) is many times the mean steady-state value: \( \langle n \rangle = ab \).
High $\gamma$ implies bursts of protein synthesis

Figure 5: As $\gamma$ increases, protein synthesis occurs in bursts. Time courses of protein numbers from simulations of the two-stage model of of Fig. 1. When $\gamma$ is increased to 100 from 1, we see steep bursts of synthesis: short-lived mRNAs are only able to be occasionally translated before being degraded. The protein degradation rate is $d_1 = 0.0005s^{-1}$.  

\[ a = 20 \text{ and } b = 2.5 \]  

\[ a = 0.5 \text{ and } b = 100 \]  

Both examples have a mean protein number of 50.

Solving the master equation for bursts of protein synthesis

When $\gamma \gg 1$, the distribution for protein numbers can also be derived by only considering $P_n(\tau)$, the probability of having $n$ proteins at time $\tau$, if this probability obeys a master equation where proteins are synthesized in bursts. We let the size $r$ of a burst obey a geometric distribution,

\[ P(r) = \left( \frac{b}{1+b} \right)^r \left( 1 - \frac{b}{1+b} \right). \]  

(53)

The corresponding master equation is

\[ \frac{\partial P_n}{\partial \tau} = a \left[ \left( 1 - \frac{b}{1+b} \right) \sum_{r=0}^{n} \left( \frac{b}{1+b} \right)^r P_{n-r} - P_n \right] + (n+1)P_{n+1} -nP_n \]  

(54)
which can be converted into an equation for the generating function, \( F(z) = \sum_n z^n P_n(\tau) \).

The generating function obeys

\[
\frac{\partial F}{\partial \tau} = (1 - z) \frac{\partial F}{\partial z} - aF + a \left( 1 - \frac{b}{1 + b} \right) \sum_{n=0}^{\infty} \sum_{r=0}^{n} z^n \left( \frac{b}{1 + b} \right)^r P_{n-r}
\]  

(55)

where we need to evaluate the sums over \( n \) and \( r \). Relabelling and resuming

\[
\sum_{n=0}^{\infty} \sum_{r=0}^{n} z^n \left( \frac{b}{1 + b} \right)^r P_{n-r} = \sum_{n=0}^{\infty} \sum_{k=0}^{n} z^n \left( \frac{b}{1 + b} \right)^{n-k} P_k
\]

\[
= \sum_{k=0}^{\infty} \left( \frac{b}{1 + b} \right)^{-k} P_k \sum_{n=k}^{\infty} \left( \frac{b}{1 + b} \right)^n
\]

\[
= \sum_{k=0}^{\infty} \frac{P_k \left( \frac{bz}{1 + b} \right)^k}{\left( 1 - \frac{bz}{1 + b} \right) \left( \frac{b}{1 + b} \right)^k}
\]

\[
= \frac{F(z)}{1 - \frac{bz}{1 + b}}
\]  

(56)

where we use the definition of the generating function. Consequently, Eq. 55 becomes

\[
\frac{\partial F}{\partial \tau} = (1 - z) \frac{\partial F}{\partial z} + \left( 1 - \frac{b}{1 + b} \right) - 1 \right) aF
\]  

(57)

or

\[
\frac{1}{v} \frac{\partial F}{\partial \tau} + \frac{\partial F}{\partial v} = \frac{ab}{1 - bv} F
\]  

(58)

with \( v = z - 1 \). This partial differential equation is Eq. 23 when \( \gamma \gg 1 \) and Eq. 31 holds.

### Derivation of the gamma distribution for protein numbers

We can derive the gamma distribution for protein numbers found by Friedman et al. [5] when \( n \) is large. If \( P(n|a, b) \) is the negative binomial distribution and \( \Gamma(n|a, b) \) is the gamma distribution, then

\[
P(n|a, b) = \int_0^\infty d\lambda \frac{e^{-\lambda} \lambda^n}{n!} \Gamma(\lambda|a, b)
\]  

(59)

which is a general relation between the negative binomial and gamma distributions. It can be verified by evaluating the integral using the definition of a gamma function [2]. If we approximate the Poisson distribution by a normal distribution and write \( z = \lambda - n \), Eq. 59 becomes

\[
P(n|a, b) \approx \int_{-\infty}^{\infty} dz \frac{e^{-\frac{z^2}{2(z+n)}}}{\sqrt{2\pi(z+n)}} \Gamma(z + n|a, b)
\]

\[
= \int_{-\infty}^{\infty} dz \frac{e^{-\frac{z^2}{2(1+z)^n}}}{\sqrt{2\pi n}} \cdot \left( 1 + \frac{z}{n} \right)^{-\frac{1}{2}} \Gamma\left(n \left[1 + \frac{z}{n}\right]|a, b\right).
\]  

(60)
We note that only values of \( z \) close to zero contribute to the integral when \( n \gg 1 \) because \( z = 0 \) is the minimum of the exponent in the integrand. Then \( n \gg 1 \) implies \( z/n \ll 1 \), and so

\[
P(n|a,b) \simeq \int_{-\infty}^{\infty} dz \frac{e^{-\frac{z^2}{2n}}}{\sqrt{2\pi n}} \Gamma(n|a,b) = \Gamma(n|a,b)
\]

for large \( n \), as expected [5].

### Derivation of the protein distribution for a three-stage model of gene expression

We can use the same approximation of large \( \gamma \) to find the protein distribution for the three-stage model. Let \( P_{m,n}^{(0)} \) be the probability of having \( m \) mRNAs and \( n \) proteins when the DNA is inactive and \( P_{m,n}^{(1)} \) be the probability of having \( m \) mRNAs and \( n \) proteins when the DNA is active. The master equation consists of two coupled equations:

\[
\frac{\partial P_{m,n}^{(0)}}{\partial \tau} = \kappa_1 P_{m,n}^{(1)} - \kappa_0 P_{m,n}^{(0)} + (n+1)P_{m,n+1}^{(0)} -nP_{m,n},
\]

\[
\frac{\partial P_{m,n}^{(1)}}{\partial \tau} = -\kappa_1 P_{m,n}^{(1)} + \kappa_0 P_{m,n}^{(0)} + (n+1)P_{m,n+1}^{(1)} -nP_{m,n}^{(1)} + a \left( P_{m-1,n}^{(1)} - P_{m,n}^{(1)} \right)
\]

\[
\frac{\partial P_{m,n}^{(0)}}{\partial \tau} = \gamma \left[ (m+1)P_{m+1,n}^{(0)} - mP_{m,n}^{(0)} + b(m,n) \right]
\]

\[
\frac{\partial P_{m,n}^{(1)}}{\partial \tau} = \gamma \left[ (m+1)P_{m+1,n}^{(1)} - mP_{m,n}^{(1)} + b(m,n) \right]
\]

where \( \kappa_0 = k_0/d_1 \) and \( \kappa_1 = k_1/d_1 \). By defining two generating functions

\[
f^{(0)}(z',z) = \sum_{m,n} (z')^m z^n P_{m,n}^{(0)} ;
\]

\[
f^{(1)}(z',z) = \sum_{m,n} (z')^m z^n P_{m,n}^{(1)}
\]

these equations become

\[
\frac{1}{v} \frac{\partial f^{(0)}}{\partial \tau} = \frac{1}{v} \left[ \kappa_1 f^{(1)} - \kappa_0 f^{(0)} \right] - \frac{\partial f^{(0)}}{\partial u} + \gamma \left[ b(1+u) - \frac{u}{v} \right] \frac{\partial f^{(0)}}{\partial u}
\]

\[
\frac{1}{v} \frac{\partial f^{(1)}}{\partial \tau} = \frac{1}{v} \left[ \kappa_1 f^{(1)} + \kappa_0 f^{(0)} \right] - \frac{\partial f^{(1)}}{\partial u} + \frac{u}{v} f^{(1)} + \gamma \left[ b(1+u) - \frac{u}{v} \right] \frac{\partial f^{(1)}}{\partial u}
\]

with \( u = z' - 1 \) and \( v = z - 1 \).

At steady-state \( \frac{\partial f^{(0)}}{\partial \tau} = \frac{\partial f^{(1)}}{\partial \tau} = 0 \), and we find using the method of characteristics that

\[
\frac{dv}{dr} = 1 ;
\]

\[
\frac{df^{(0)}}{dr} = \frac{1}{v} \left[ \kappa_1 f^{(1)} - \kappa_0 f^{(0)} \right] ;
\]

\[
\frac{du}{dr} = \frac{1}{v} \left[ -\kappa_1 f^{(1)} + \kappa_0 f^{(0)} + a \frac{u}{v} f^{(1)} \right]
\]

where \( r \) measures the distance along a characteristic. Both \( u \) and \( v \) obey Eq. 24 again. Consequently, \( v = r \) and \( u \simeq \frac{bv}{1-bv} \) from Eq. 31 when \( \gamma \gg 1 \). From Eq. 67, we therefore obtain the two coupled differential equations:

\[
\frac{df^{(0)}}{dv} = \kappa_1 f^{(1)} - \kappa_0 f^{(0)}
\]

\[
\frac{df^{(1)}}{dv} = -\kappa_1 f^{(1)} + \kappa_0 f^{(0)} + \frac{abv}{1-bv} f^{(1)}.
\]
Following Hornos et al. [6], Eqs. 68 and 69 can be reduced to one differential equation for \( f^{(0)}(v) \) by solving Eq. 68 for \( f^{(1)} \) in terms of \( f^{(0)} \) and its derivative, and inserting the result into Eq. 69. This equation becomes a second-order differential equation:

\[
v(bv - 1) \frac{df^{(0)}}{dv^2} + [(\kappa_0 + \kappa_1)(bv - 1) + bv(1+a) - 1] \frac{df^{(0)}}{dv} + ab\kappa_0 f^{(0)} = 0. \tag{70}
\]

Eq. 70 has solution

\[
f^{(0)}(v) = C \, 2F_1(\alpha, \beta, 1 - \kappa_0 - \kappa_1; bv) \tag{71}
\]

where \( 2F_1(a, b, c; z) \) is a hypergeometric function,

\[
\alpha = \frac{1}{2} \left( a + \kappa_0 + \kappa_1 + \sqrt{(a + \kappa_0 + \kappa_1)^2 - 4a\kappa_0} \right) \tag{72}
\]

\[
\beta = \frac{1}{2} \left( a + \kappa_0 + \kappa_1 - \sqrt{(a + \kappa_0 + \kappa_1)^2 - 4a\kappa_0} \right), \tag{73}
\]

and \( C \) is a constant of integration.

We can find the generating function for protein numbers, \( F(z) = f^{(0)}(z) + f^{(1)}(z) \), by using our solution for \( f^{(0)} \) and Eq. 68 to find \( f^{(1)} \). Determining the constant of integration \( C \) from \( F(1) = 1 \) and using the relation \( c(c+1) \, 2F_1(a, b, c; z) = c(c+1) \, 2F_1(a, b, c+1; z) + abz \, 2F_1(a+1, b+1, c+2; z) \), we find that

\[
F(z) = 2F_1(\alpha, \beta, \kappa_0 + \kappa_1; b(z - 1)), \tag{74}
\]

replacing \( v \) by \( z - 1 \).

Expanding the generating function around \( z = 0 \) determines the probabilities \( P_n \). Using properties of the \( n \)-th derivatives with respect to \( z \) of the hypergeometric function, \( 2F_1^{(n)}(a, b, c; z) \), we can write

\[
F(z) = \sum_{n=0}^{\infty} \frac{2F_1^{(n)}(\alpha, \beta, \kappa_0 + \kappa_1; -b)}{n!} \frac{b^n}{n!} z^n
= \sum_{n=0}^{\infty} \frac{\Gamma(\alpha + n)\Gamma(\beta + n)\Gamma(\kappa_0 + \kappa_1) b^n}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\kappa_0 + \kappa_1 + n)n!} \, 2F_1(\alpha + n, \beta + n, \kappa_0 + \kappa_1 + n; -b) z^n \tag{75}
\]

and \( P_n \) can be found from the definition of \( F(z) \): \( F(z) = \sum_n P_n z^n \). With the linear transformation formulæ for hypergeometric functions [2], we write \( P_n \) as

\[
P_n = \frac{\Gamma(\alpha + n)\Gamma(\beta + n)\Gamma(\kappa_0 + \kappa_1)}{\Gamma(n + 1)\Gamma(\alpha)\Gamma(\beta)\Gamma(\kappa_0 + \kappa_1 + n)} \left( \frac{b}{1+b} \right)^n \left( 1 - \frac{b}{1+b} \right) \alpha
\times 2F_1(\alpha + n, \kappa_0 + \kappa_1 - \beta, \kappa_0 + \kappa_1 + n; \frac{b}{1+b}). \tag{76}
\]

**The exact mRNA distributions**

For completeness, we include the mRNA distributions for the two-stage and three-stage models. With initially zero mRNAs, the two-stage model has a Poisson distribution:

\[
P_m(t) = e^{-\langle m(t) \rangle} \frac{\langle m(t) \rangle^m}{m!} \tag{77}
\]
where $\langle m(t) \rangle = m_s \left(1 - e^{-d_0 t}\right)$ and $m_s = v_0/d_0$ is the steady-state number of mRNAs. The propagator probability satisfies

$$P_{m|k}(t) = \sum_{r=0}^{k} \binom{k}{r} P_{m-r}(t) \left(1 - e^{-d_0 t}\right)^{k-r} e^{-rd_0 t}$$

(78)

with $P_n(t) = 0$ if $m < 0$.

The steady-state distribution of mRNA for the three-stage model was first derived by Peccoud and Ycart, although they did not recognize it as such [7], and also by Raj et al. [8]. The exact probability of having $m$ RNAs at steady-state is

$$P_m = \frac{m_s^m e^{-m_s}}{m!} \cdot \frac{\Gamma(\zeta_0 + m)\Gamma(\zeta_0 + \zeta_1)}{\Gamma(\zeta_0 + \zeta_1 + m)\Gamma(\zeta_0)} \, _1F_1(\zeta_1, \zeta_0 + \zeta_1 + m; m_s)$$

(79)

where $m_s = v_0/d_0$, $\zeta_0 = k_0/d_0$, and $\zeta_1 = k_1/d_0$, and $_1F_1(a, b; z)$ is the confluent hypergeometric function of the first kind [2]. Eq. 79 like Eq. 18 can be bimodal. For $\zeta_1 = k_1/d_0 \gg 1$, Eq. 79 tends to a negative binomial distribution [8], because then mRNA synthesis is more burst-like. The distribution becomes Poisson when $k_1$ is zero, and the three-stage model reduces to the two-stage model.

References


