

Stochastic Dynamics of RNA interference: A Single-Molecule Time-based Approach

Baihan Lin^{1*}

¹Department of Applied Mathematics, University of Washington, Seattle, WA 98195, USA

*doerlbh@gmail.com

ABSTRACT

Under a general mechanistic understanding of RNA interference, we present a stochastic model for the cellular dynamics of gene expression involved in RNA interference in a single-molecule time-based approach. Although the mechanistic model is constructed under various simplifications and assumptions to decrease the innate diversity and complexity in RNA interference behaviors, considering the introduction of single-molecule approach introduced into this system, the biological implication is still a worthwhile effort. Both the deterministic model and stochastic model demonstrate similar equilibrium. A Monte Carlo simulation method is also introduced to support chemical master equation. Although this is not the first paper to establish a stochastic theory in RNA level, the understanding of stochastic behavior and how it compares with the traditional deterministic approach still offers certain insights into a non-closed biochemical system.

Introduction

RNA interference (RNAi) is an important biological process in which RNA molecules inhibit gene expression. Although in different mechanisms, typically these noncoding RNA molecules causes the destruction of specific messenger RNA (mRNA) molecules.¹ As shown in Figure 1, the RNA interference can be simplified as a genetic circuit, where central dogma determines that one set of DNA was transcribed into messenger RNA (mRNA) and interfering RNA (iRNA), mRNA were translated to protein, while iRNA inhibit the translation of mRNA.

Inspired by the single-molecule approach applied to probe the stochastic gene expression with respect to transcription factor² and stochastic gene state switching³, we proposed a single-molecule theory of stochastic gene expression on RNA-based regulation, starting from interference system of siRNA and microRNA which reversibly inhibits mRNA translation with increased degradation.

microRNA (miRNA) and small interfering RNA (siRNA) two types of small RNA molecules central to RNA interference. Take microRNA as an example: a protein called exportin-5 transports a hairpin primary microRNA (pri-miRNA) out of nucleus; an enzyme called dicer trimes pri-miRNA and removes hairpin loop, exposing one strand to join a group of proteins forming microRNA-protein complex; the microRNA-protein complex binds with mRNA strand, to produce a RNA-induced silencing complex (RISC), which stops translation process; the prescence of microRNA-protein complex blocks translation as well as speeding up deadenylation, which is a process to breakdown Poly-A tail and cause mRNA to be degraded sooner and translated less. siRNA's regulatory pathway is similar to microRNA.

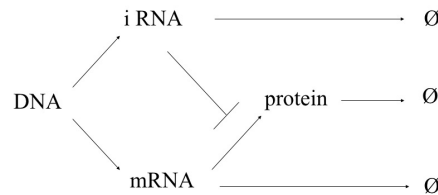


Figure 1. a diagram for the genetic circuit simplification of RNA interference

Results

To simply the model, we used iRNA to represent microRNA and siRNA, which are similar RNA interference agents. Instead of producing transcription factor to bind with DNA, iRNA tends to bind with mRNA and form a complex to inhibit the translation.

The chemical reactions involved in this process is therefore summarized as these 8 steps:

- $\text{DNA} + \text{NTPs} \xrightarrow{k_1} \text{DNA} + \text{mRNA}$
- $\text{DNA} + \text{NTPs} \xrightarrow{k_2} \text{DNA} + \text{iRNA}$
- $\text{mRNA} + \text{AAs} \xrightarrow{k_3} \text{mRNA} + \text{protein}$
- $\text{mRNA} + \text{iRNA} + \text{RISC} \xrightleftharpoons[\text{f}]{\text{h}} \text{mRNA} \cdot \text{iRNA} \cdot \text{RISC}$
- $\text{mRNA} \xrightarrow{\gamma^1} \emptyset$
- $\text{iRNA} \xrightarrow{\gamma^1} \emptyset$
- $\text{mRNA} \cdot \text{iRNA} \cdot \text{RISC} \xrightarrow{\gamma^2} \emptyset$
- $\text{protein} \xrightarrow{\gamma^3} \emptyset$

where NTPs stand for nucleoside triphosphates, which are building blocks for RNA, and AAs stand for amino acid, which are building blocks for protein.

Deterministic Dynamics and Nonequilibrium Steady States

To simplify our model, we made several assumptions:

- we assume same degradation rate for iRNA and mRNA;
- we assume an irreversible inhibition, which means $h \gg f$;
- we assume the degradation rate of binded mRNA higher than mRNA and iRNA;
- we assume the concentrations of DNA, AAs, NTPs, RISCs are abundant and can be considered to be maintained.
- we assume no other regulation behavior (such as transcription factors etc.)

Based on above reactions, following the law of mass action, the ordinary differential equation can be written:

$$\frac{d[\text{iRNA}]}{dt} = k_2[\text{DNA}][\text{NTPs}] - \gamma_1[\text{iRNA}] + f[\text{mRNA} \cdot \text{iRNA} \cdot \text{RISC}] - h[\text{RISCs}][\text{iRNA}][\text{mRNA}] \quad (1)$$

$$\frac{d[\text{mRNA}]}{dt} = k_1[\text{DNA}][\text{NTPs}] - \gamma_1[\text{mRNA}] + f[\text{mRNA} \cdot \text{iRNA} \cdot \text{RISC}] - h[\text{RISCs}][\text{iRNA}][\text{mRNA}] \quad (2)$$

$$\frac{d[\text{mRNA} \cdot \text{iRNA} \cdot \text{RISC}]}{dt} = -\gamma_2[\text{mRNA} \cdot \text{iRNA} \cdot \text{RISC}] + h[\text{RISCs}][\text{iRNA}][\text{mRNA}] \quad (3)$$

$$\frac{d[\text{protein}]}{dt} = k_3[\text{mRNA}][\text{NTPs}] - \gamma_3[\text{protein}] \quad (4)$$

Since the concentrations of DNA, AAs, NTPs, RISCs can be considered to be maintained in the cell, we can simplify the system by denoting:

$$K_1 = k_1[\text{DNA}][\text{NTPs}], K_2 = k_2[\text{DNA}][\text{NTPs}], K_3 = k_3[\text{AAs}], H = h[\text{RISCs}] \quad (5)$$

If we denote the concentrations of iRNA, mRNA, mRNA.iRNA.RISC and protein at time t, denoted as x(t), y(t), z(t), w(t), then the differential equations can be rewritten as:

$$\frac{dx}{dt} = K_2 - \gamma_1 x + fz - Hxy \quad (6)$$

$$\frac{dy}{dt} = K_1 - \gamma_1 y + fz - Hxy \quad (7)$$

$$\frac{dz}{dt} = -\gamma_2 z + Hxy \quad (8)$$

$$\frac{dw}{dt} = K_3 y - \gamma_3 w \quad (9)$$

If we nondimensionalize all the variables by setting

$$v = \frac{\gamma_1}{K_2} x, r = \frac{\gamma_1}{K_1} y, q = \gamma_2 H z, s = \frac{\gamma_1 \gamma_3}{K_1 K_3} w, \tau = \gamma_3 t. \quad (10)$$

We get:

$$\frac{dv}{d\tau} = \varepsilon_1 (1 - v - H_1 vr) = f(v, r) \quad (11)$$

$$\frac{dr}{d\tau} = \varepsilon_1 (1 - r - H_2 vr) = g(v, r) \quad (12)$$

$$\frac{dq}{d\tau} = \varepsilon_2 (-q + H_3 vr) \quad (13)$$

$$\frac{ds}{d\tau} = r - s \quad (14)$$

where

$$\varepsilon_1 = \frac{\gamma_2}{\gamma_3}, \varepsilon_2 = \frac{\gamma_1}{\gamma_3}, H_1 = H \frac{K_1}{\gamma_1 \gamma_3}, H_2 = H \frac{K_2}{\gamma_1 \gamma_3}, H_3 = \varepsilon_1 H^2 \frac{K_1 K_2}{\gamma_1^2}. \quad (15)$$

To obtain steady states, we set $f(v^*, r^*) = g(v^*, r^*) = 0$:

$$v^* = \frac{1}{1 + H_1 r^*} \quad (16)$$

$$r^* = \frac{1}{1 + H_2 v^*} = \frac{1}{1 + H_2 v^*} = \frac{1}{1 + H_2 \frac{1}{1 + H_1 r^*}} = \frac{1 + H_1 r^*}{1 + H_2 v^* + H_2} \quad (17)$$

Solving this quadratic equation, we get the analytical steady states as:

$$v^* = \frac{H_2 - H_1 - 1 + \sqrt{(H_2 - H_1 - 1)^2 + 4H_2}}{2H_2} \quad (18)$$

$$r^* = \frac{H_1 - H_2 - 1 + \sqrt{(H_1 - H_2 - 1)^2 + 4H_1}}{2H_1} \quad (19)$$

in which there is only one real positive steady state, and the protein production correspondent have a steady state of

$$s^* = r^* + (s(0) - r^*) e^{-\tau} \quad (20)$$

There is no bifurcation behavior in this deterministic system.

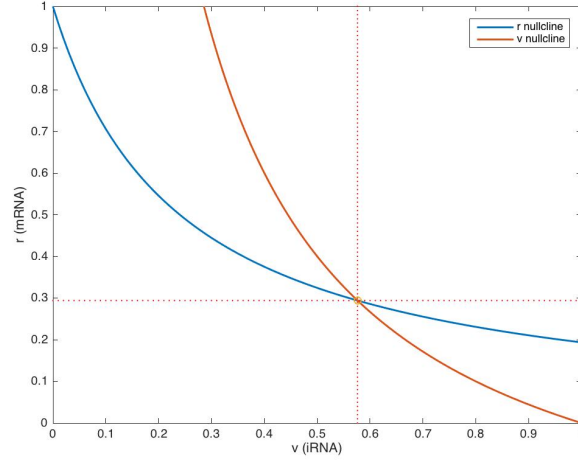


Figure 2. the phase portrait of the nondimensionalized system derives one real positive steady state calculated based on the nullclines.

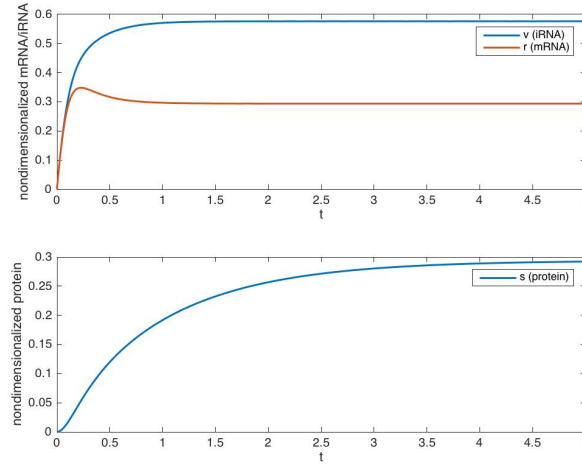


Figure 3. the deterministic dynamics of the system of mRNA, iRNA and protein.

Chemical Master Equation and Single-Molecule Kinetics

The chemical master equation of this system can be specified as:

$$\begin{aligned}
 \frac{dP(m,n,p)}{dt} = & K_2 P(m,n-1,p) + K_1 P(m-1,n,p) + mK_3 P(m,n,p-1) \\
 & + [(n+1)\gamma_1 + m(n+1)H] P(m,n+1,p) \\
 & + (p+1)\gamma_3 P(m,n,p+1) + [(m+1)\gamma_1 + (m+1)nH] P(m+1,n,p) \\
 & - P(m,n,p)(K_2 + P\gamma_3 + K_1 + mK_3 + n\gamma_1 + mnH + m\gamma_1 + mnH)
 \end{aligned} \tag{21}$$

$$\begin{aligned}
 \frac{dP(0,n,p)}{dt} = & K_2 P(0,n-1,p) + (n+1)\gamma_1 P(0,n+1,p) + (p+1)\gamma_3 P(0,n,p+1) \\
 & + (\gamma_1 + nH) P(1,n,p) - P(0,n,p)(K_2 + P\gamma_3 + K_1 + n\gamma_1)
 \end{aligned} \tag{22}$$

$$\begin{aligned} \frac{dP(m,0,p)}{dt} = & K_1 P(m-1,0,p) + mK_3 P(m,0,p-1) + [\gamma_1 + mH] P(m,1,p) \\ & + (p+1)\gamma_3 P(m,0,p+1) + (m+1)\gamma_1 P(m+1,0,p) \\ & - P(m,0,p)(K_2 + P\gamma_3 + K_1 + mK_3 + m\gamma_1) \end{aligned} \quad (23)$$

$$\begin{aligned} \frac{dP(m,n,0)}{dt} = & K_2 P(m,n-1,-) + K_1 P(m-1,n,0) + [(n+1)\gamma_1 + m(n+1)H] P(m,n+1,0) \\ & + \gamma_3 P(m,n,1) + [(m+1)\gamma_1 + (m+1)nH] P(m+1,n,0) \\ & - P(m,n,0)(K_2 + P\gamma_3 + K_1 + mK_3 \\ & + n\gamma_1 + mnH + m\gamma_1 + mnH) \end{aligned} \quad (24)$$

As shown in Figure 4, the transition diagram is in 3 dimensions, in the number of iRNA molecule, mRNA molecule, and protein molecule. Different from the stochastic system of transcription factor, the iRNA and mRNA axes are independent of protein axis.

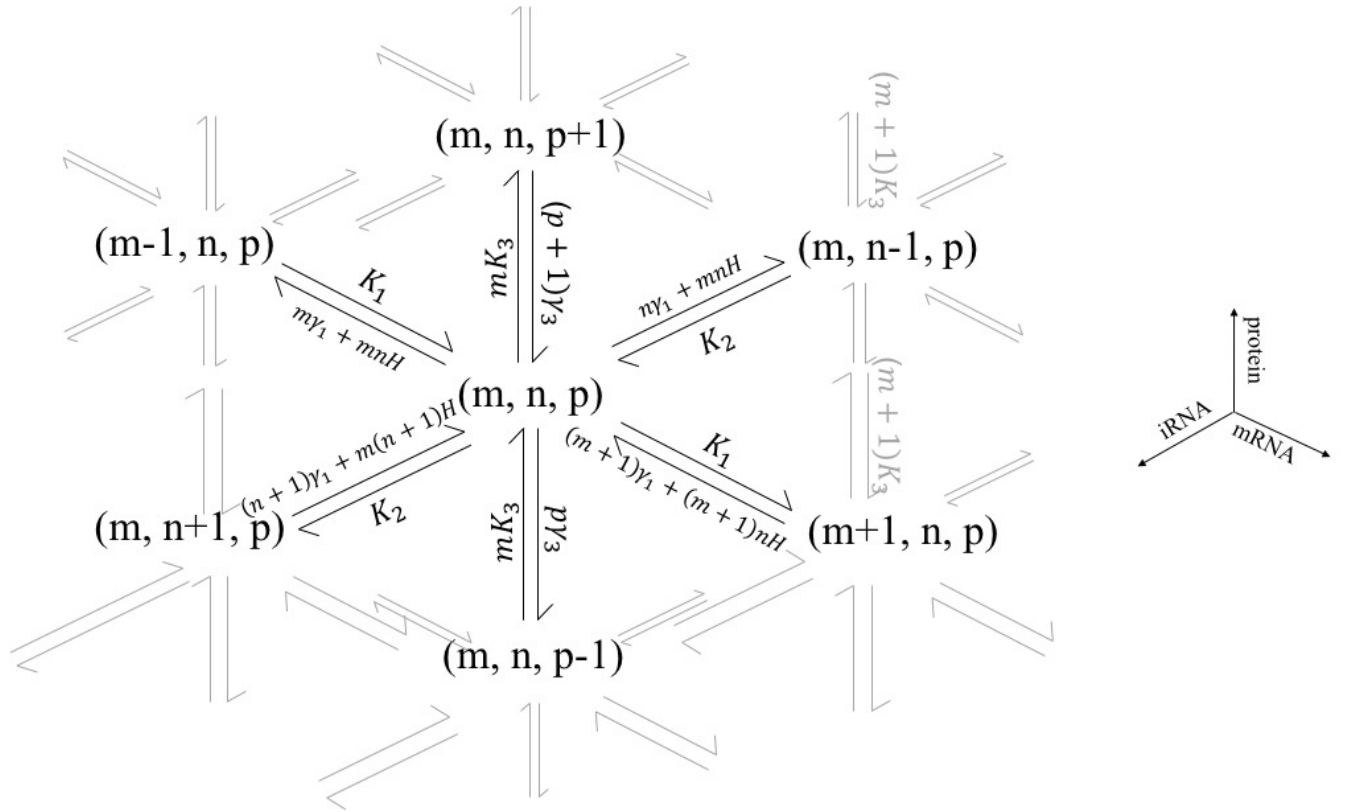


Figure 4. a discrete schematic illustrating the Markovian kinetics of a single molecule of iRNA, mRNA, or protein with conformational fluctuations.

However, without accessible experimental data from single-molecule measurement of RNA interference to provide insights into the probability distribution of the system with respect to mRNA, iRNA and protein, we cannot produce probabilities generating functions to analytically solve the entire system.

Therefore, we took a different approach towards this system. If we consider all the reactions involved in this system (synthesis, transcription, regulation, translation and degradation) as independent reactions, then under the single-molecule level, within a high resolution of measurement, only one reaction can happen at the same time. Thus, we can assume a Poisson distribution of the waiting time until next reaction

For each reaction type, the mean time till the next reaction occurs again is the inverse of the reaction rate. Since the rate at which anything will happen is simply the sum of all the reaction rates, which we'll denote by λ as shown as Figure 5, we also can write the mean time until any reaction occurs as

```

rates = [ K1;
          K2;
          K3;
          H;
          gamma1*y;
          gamma1*x;
          gamma2*z;
          gamma3*w ];

```

Figure 5. the reaction rates for each reaction.

$$\tau_m = \frac{1}{\lambda}$$

Then, under the assumption of a Poisson process, it's fairly straightforward to pick the next most likely reaction and the amount of time until the reaction occurs. For the time, we have

$$\tau = \frac{1}{\lambda} \log\left(\frac{1}{U_1}\right)$$

for U_1 drawn uniformly from the unit interval. Likewise, if we denote the individual reaction rates as a_i we can select one based on the relative reaction rates by choosing a random variable

$$r = \lambda * U_2$$

which gives for U_2 drawn uniformly from the unit interval. This gives us a random variable in the range $(0, \lambda)$ which we can use to select the next reaction j by finding j such that

$$\sum_0^{j-1} \leq r \leq \sum_0^j$$

We then update the current time step with that selected reaction from the transition matrix (Figure 6), the number of reactants in all compartments, and go back to the beginning. The update steps are performed by keeping matrices to store the reactant/reaction combinations as well as the number of molecules used in each reaction. Implementation details can be found in the supplementary information.

```

transition =
    0     1     0    -1     0    -1     0     0
    1     0     0    -1    -1     0     0     0
    0     0     0     1     0     0    -1     0
    0     0     1     0     0     0     0    -1

```

Figure 6. the transition matrix for this stochastic system.

For each reaction type, the mean time till the next reaction occurs again is the inverse of the reaction rate. Since the rate at which anything will happen is simply the sum of all the reaction rates, which we'll denote by λ , we also can write the mean time until any reaction occurs as

As shown in Figure 7, the stochastic behavior is well captured in this Monte Carlos simulation in considerable amount of run time. Similar to the specified chemical master equation, the stochastic relationship of the system is shown in Figure 8 , starting from (0, 0, 0), departing towards an equilibrium region.

Discussion

As shown from our investigation, the system tends to reach an equilibrium steady state within certain amount of time. The chemical master equation captures the single-molecule event in a precise level (Figure 3). However, the lack of experimental

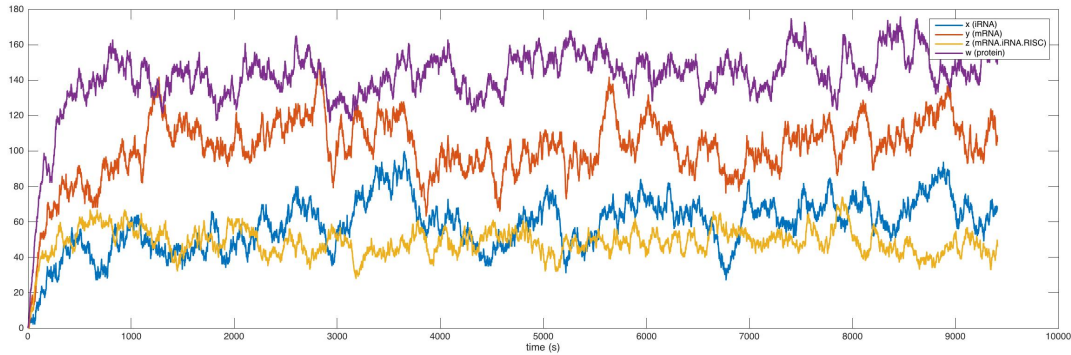


Figure 7. the stochastic dynamics of the system of mRNA, iRNA and protein along time.

understanding of the probability distribution of the mRNA, iRNA and protein in a single-molecule level prevent us from developing probability generating functions, which can give us a better understanding of the whole picture. Despite this infeasible attempt, another attempt to introduce single-molecule stochastic dynamics based on a waiting time model was well constructed under the assumption of a Poisson distribution. The Monte Carlo simulation based on this stochastic model captured this behavior and demonstrated the predicted equilibrium behavior (Figure 7). Figure 8 is also a nice representation of how our chemical master equation can be simulated with more probability generating information.

Based on our model, more detailed investigation can be conducted. Specific types of microRNA can have cascade-like genetic circuits as well as specific binded RNA sequences, which implies combinatorial structures of signaling. We can also consider an introduction of transcription factor, in order to observe the interactions between these two types of regulations. Another improvement would be reducing our assumptions one by one.

Further exploration can also be expanded to the stochastic modeling of lncRNA's "origami-like" interaction with DNA and RNA in a 3D single-molecule biophysics perspective, which can be very beneficial to future bioengineering regarding disease and development. With the same genomic code, the massive cellular diversity might be just encoded by regulation of different noncoding RNA. Some promising future attempt would be including the noncommutative sequential regulation in which a time-evolving origami can generate diversity in biological systems.

References

1. MacRae, I. J. *et al.* Structural basis for double-stranded RNA processing by Dicer. *Science* **311**, 195–198 (2006). URL <http://science.sciencemag.org/content/311/5758/195>. <http://science.sciencemag.org/content/311/5758/195.full.pdf>.
2. Elf, J., Li, G.-W. & Xie, X. S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191–1194 (2007). URL <http://science.sciencemag.org/content/316/5828/1191>. <http://science.sciencemag.org/content/316/5828/1191.full.pdf>.
3. Ge, H., Qian, H. & Xie, X. S. Stochastic phenotype transition of a single cell in an intermediate region of gene state switching. *Phys. Rev. Lett.* **114**, 078101 (2015). URL <http://link.aps.org/doi/10.1103/PhysRevLett.114.078101>.

Acknowledgements

Thank Prof. Qian for giving us insightful lectures about the mathematical theories of cellular dynamics and the exciting field of mathematical biology. Thank him for always setting challenging questions for me to explore! Thank my friends in AMATH 531 who come along with me in great energy! Thank University of Washington for giving me the platform to scientifically explore problems and subjects! I will continue the voyage of exploring the infinite realm of mathematical and systems biology in my academic career fearlessly.

Additional information

For supporting MATLAB Codes, please refer to Supplementary Information Attached. All code for the reproduction of the reported results can be downloaded from my [GitHub Repository](#).

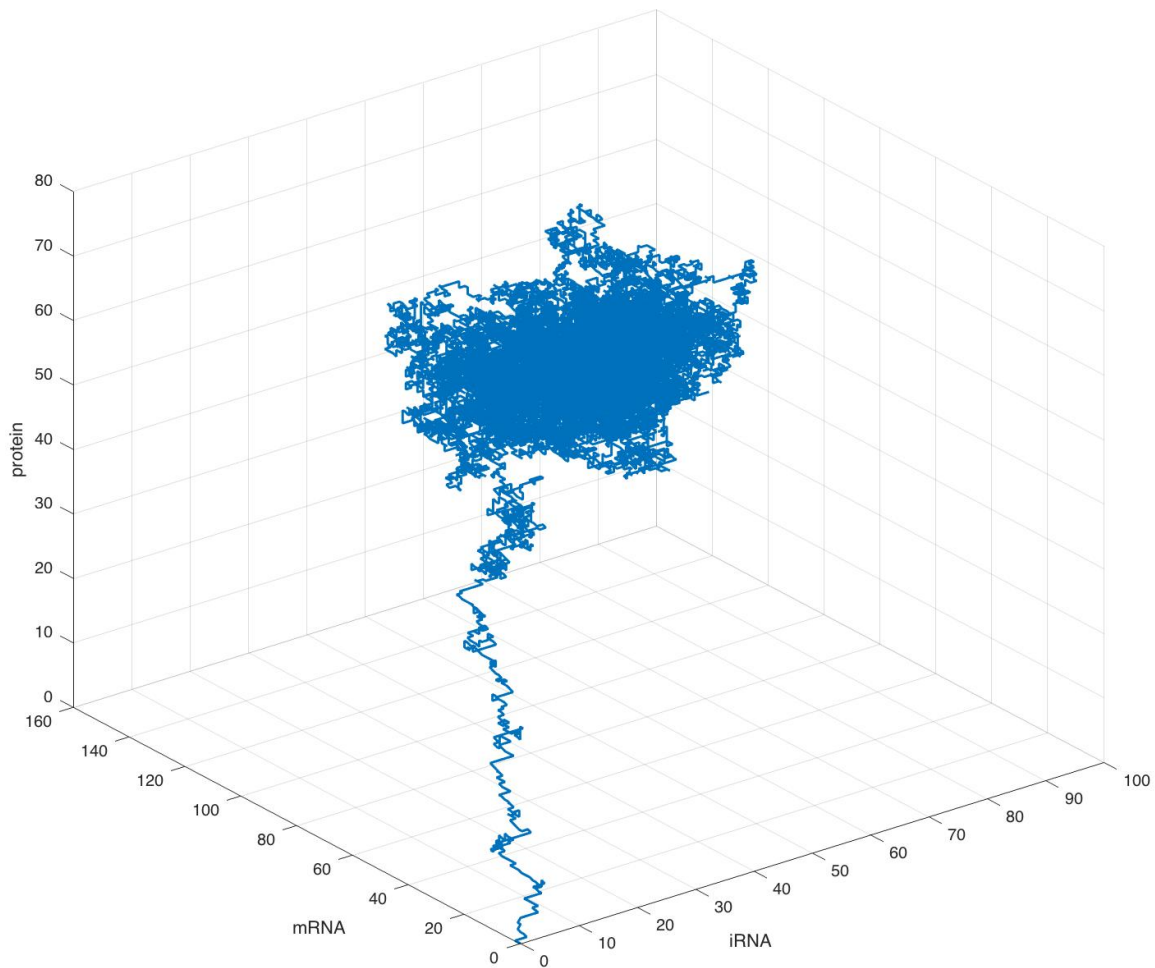


Figure 8. the stochastic relationship of mRNA, iRNA and protein along time.

Supplementary information

Supporting MATLAB codes are attached below:

```
% RNA interference modeled by Deterministic ODE
% Author: Baihan Lin
% Date: Nov 2016

clear all; close all

gm1 = 0.3;
gm2 = 0.8;
gm3 = 0.2;
K1 = 0.3;
K2 = 0.5;
H = 0.5;

ep1 = gm2/gm3;
```



```

ep2 = gm1/gm3;
H1 = H*K1/(gm1*gm3);
H2 = H*K2/(gm1*gm3);
H3 = H^2*ep1*K1*K2/(gm2^2);

%% Dynamic Simulation

for numreps=1:1

p = [ep1, ep2, H1, H2, H3];
x0 = [0, 0, 0, 0];

Tmax=5;
options = odeset('NonNegative',1:4); %make solutions nonnegative

[T,Y] = ode45(@(t,y) iRNA_ODE(t,y,p),[0 Tmax],x0,options);

figure(1)
subplot(2,1,1);
plot(T,Y(:,1:2),'LineWidth',1.5);
legend('v (iRNA)', 'r (mRNA)')
xlabel('t');
ylabel('nondimensionalized mRNA/iRNA');

subplot(2,1,2);
plot(T,Y(:,4),'LineWidth',1.5); hold on;
legend('s (protein)')
xlabel('t');
ylabel('nondimensionalized protein');

figure(2)
grid on;
set(gca,'FontSize',16)
plot3(Y(:,1),Y(:,2),Y(:,4),'LineWidth',1.5);
xlabel('v (iRNA)');
ylabel('r (mRNA)');
zlabel('s (protein)');

end

%% Steady States Search

v = 0:0.01:1;
r = 0:0.01:1;
vpre = 1+H1*r;
rpre = 1+H2*v;
vnull = 1./vpre;
rnull = 1./rpre;

v_star = (H2-H1-1+sqrt((H2-H1-1)^2+4*H2))/(2*H2);
r_star = (H1-H2-1+sqrt((H1-H2-1)^2+4*H1))/(2*H1);

figure(3)
plot(v, rnull,'LineWidth',1.5);hold on
plot(vnull,r,'LineWidth',1.5);
plot(v, r_star*ones(1,length(v)), 'r');
plot(v_star*ones(1,length(v)), r, 'r');
plot(v_star,r_star,'o');
legend('r nullcline', 'v nullcline');
xlabel('v (iRNA)');
ylabel('r (mRNA)');

% RNA interference modeled by Deterministic ODE
% Author: Baihan Lin
% Date: Nov 2016

% dvdt = ep1 * (1 - v - H1 * v * r);

```

```

% drdt = ep1 * (1 - r - H2 * v * r);
% dqdt = ep2 * (- q + H3 * v * r);
% dsdt = r - s;

function dy = iRNA_ODE(t,y,p)

ep1 = p(1);
ep2 = p(2);
H1 = p(3);
H2 = p(4);
H3 = p(5);

dy = zeros(4,1);

dy(1) = ep1 * (1 - y(1) - H1 * y(1) * y(2));
dy(2) = ep1 * (1 - y(2) - H2 * y(1) * y(2));
dy(3) = ep2 * (- y(3) + H3 * y(1) * y(2));
dy(4) = y(2) - y(4);

end

% RNA interference modeled by Monte Carlos Simulation
% Author: Baihan Lin
% Date: Nov 2016

clear all; close all

gm1 = 0.005;
gm2 = 0.01;
gm3 = 0.006;
K1 = 1;
K2 = 0.8;
K3 = 0.9;
H = 0.5;
f = 0.3;

% ep1 = gm2/gm3;
% ep2 = gm1/gm3;
% H1 = H*K1/(gm1*gm3);
% H2 = H*K2/(gm1*gm3);
% H3 = H^2*ep1*K1*K2/(gm2^2);

p = [gm1, gm2, gm3, K1, K2, K3, H, f];
x0 = [0; 0; 0; 0];

rng(1);

Nstep=50000;
stptime = zeros(Nstep,1);
time = zeros(Nstep,1);
xall = zeros(Nstep,4);
xall(1,:) = x0;
x = x0;

for step = 1 : Nstep - 1
    [xnew, tau] = stoch_update(x, p);
    x = xnew;
    stptime(step+1) = tau;
    time(step+1) = time(step) + tau;
    xall(step+1,:) = x;
end

figure(1)
plot(time,xall(:,1:4), 'LineWidth',1.5);
legend('x (iRNA)', 'y (mRNA)', 'z (mRNA.iRNA.RISC)', 'w (protein)');
xlabel('time (s)');
%
```

```

% figure(2)
% plot(1:Nstep,xall(:,1:4),'LineWidth',1.5);
% legend('x (iRNA)', 'y (mRNA)', 'z (mRNA.iRNA.RISC)', 'w (protein)');
% xlabel('steps');

figure(3)
plot3(xall(:,1),xall(:,2),xall(:,3), 'LineWidth',1.5);
grid on
xlabel('iRNA');
ylabel('mRNA');
zlabel('protein');

% RNA interference modeled by Stochastic Monte Carlos
% Author: Baihan Lin
% Date: Nov 2016

function [ynew, tau] = stoch_update(y, p)

gm1 = p(1);
gm2 = p(2);
gm3 = p(3);
K1 = p(4);
K2 = p(5);
K3 = p(6);
H = p(7);

rates = [K1;
        K2;
        K3;
        H;
        gm1*y(2);
        gm1*y(1);
        gm2*y(3);
        gm3*y(4)];

lambda = sum(abs(rates));

transition = [0 1 0 -1 0 -1 0 0;
            1 0 0 -1 -1 0 0 0;
            0 0 0 1 0 0 -1 0;
            0 0 1 0 0 0 0 -1];

ynew = y;

tau = log(1/rand()) / lambda;

r = rand()*lambda;

current = 0;
selection = 1;

for it = 1:length(rates)
    current = current + rates(it);
    if current > r
        selection = it;
        disp(selection);
        break;
    end
end

ynew
ynew = ynew + transition(:,selection);
ynew = ynew.*(ynew >= 0);

end

```