

Análisis Estocástico de Un Sistema Génico Simple para la Síntesis de una Proteína Implementando los Métodos de Gillespie

Diana Carolina Clavijo Buritica^{1,2}, Bárbara Valeria Mejía³, Lina María Rojas³, Leandro Sáenz Castro²

¹ Departamento de Biología, Universidad Nacional de Colombia. Bogotá, Colombia.

² Facultad de Medicina, Fundación Universitaria Juan N. Corpas. Bogotá, Colombia.

³ Departamento de Ingeniería Química, Universidad de los Andes. Bogotá, Colombia.

Resumen

En la mayoría de los casos en los que se requiere describir una red biológica se propone un sistema de ecuaciones diferenciales acopladas, que luego se resuelve por métodos numéricos. Sin embargo, cuando un solo valor en estado estacionario no predice el comportamiento de la población total, es indispensable que el modelo de representación describa la distribución de estados dentro un sistema. El algoritmo propuesto por Gillespie en 1998 consiste en la descripción de un fenómeno específico mediante herramientas estocásticas, en donde se predice el comportamiento de la ecuación maestra de probabilidades mediante la simulación de Montecarlo. El autor propone dos aproximaciones matemáticas para la resolución: el método de la primera reacción, y el método directo. El propósito de esta investigación fue a partir de un modelo génico simple para la síntesis de una proteína, representado en una red de Petri simple propuesta por Goss y Peccoud en 1998, comparar los resultados de dichas aproximaciones centrándose en las diferencias entre los resultados al analizar la red mediante un método determinístico clásico, y los dos métodos estocásticos de Gillespie. Finalmente realizar un análisis de sensibilidad al modelo estocástico con la prueba de hipótesis nula. Los resultados obtenidos muestran que efectivamente la población no se comporta uniformemente, por lo que es pertinente y recomendable la resolución por el método de Gillespie para este sistema y para sistemas similares; adicionalmente se corrobora que el método directo es más demandante computacionalmente y con la prueba de hipótesis nula se concluye que el número de proteínas final si se ve afectado por las variaciones en los parámetros cinéticos.

Palabras Clave: Modelos determinísticos, modelos estocásticos, método de la primera reacción, método directo

Autor de correspondencia:
Diana Clavijo
diana.calvijo@juanncorpas.edu.co

Como citar: Clavijo DC, Mejía BV, Rojas LM, Sáenz HL. Análisis Estocástico de Un Sistema Génico Simple para la Síntesis de una Proteína Implementando los Métodos de Gillespie. Revista Cuarzo 2018;24 (1) 7-16.

Recibido: 20 de abril de 2018
Aceptado: 5 de junio de 2018
Publicado: 30 de junio de 2018

Stochastic Analysis of a Simple Gene System for the Synthesis of a Protein Using the Gillespie Methods

Abstract

In most cases where it is required to describe a biological network, a system of coupled differential equations is proposed, which is then solved by numerical methods. However, when a single value in a steady state does not predict the behavior of the total population, it is essential that the representation model describes the distribution of states within a system. The algorithm proposed by Gillespie in 1998 consists of the description of a specific phenomenon using stochastic tools, where the behavior of the master probability equation is predicted by means of the Monte Carlo simulation. The author proposes two mathematical approaches to resolution: the first reaction method and the direct method. The purpose of this research was to compare the results of these approaches by focusing on the differences between the results when analyzing the network using a classical deterministic method and the two stochastic methods of Gillespie, using a simple gene model for protein synthesis, represented in a Petri simple network proposed by Goss and Peccoud in 1998. Finally, perform a sensitivity analysis to the stochastic model with the null hypothesis test. The results obtained show that the population does not behave uniformly, so it is pertinent and recommendable to use the Gillespie method for this system and similar systems; additionally, it is corroborated that the direct method is more computationally demanding, and the null hypothesis test concludes that the final number of proteins if affected by variations in kinetic parameters.

Keywords: Deterministic models, stochastic models, first reaction method, direct method

Licencia creative commons



Introducción

Con el advenimiento de las metodologías de alto rendimiento y métodos computacionales robustos, la realidad de estudiar procesos celulares a nivel de sistemas es más viable hoy en día. A medida que estos enfoques producen un gran conjunto de datos, los análisis a nivel de sistemas biológicos se están centrando en otros campos de la ciencia como la ingeniería, las matemáticas, la adaptación de enfoques computacionales y estadísticos para descifrar las relaciones entre las moléculas por medio de modelación predictiva (1,2). Es así como uno de los objetivos de la Biología de Sistemas es predecir el comportamiento dinámico de las redes genéticas y metabólicas de la célula (3).

Tradicionalmente los eventos dinámicos transcripcionales y traduccionales han sido estudiados y representados por ecuaciones diferenciales determinísticas y continuas, lo cual es válido cuando el comportamiento individual de las células es similar al comportamiento de la población (3). Sin embargo, el comportamiento dinámico de estos eventos es intrínsecamente estocástico, conduciendo a una variabilidad celular y que pueden ser estudiados con un método formal como son las redes de Petri estocásticas (SNPs) (4).

A través de los modelos determinísticos se puede describir el comportamiento dinámico de las redes biológicas y por su parte los modelos estocásticos pueden predecir fielmente la distribución de la población en estado estable (5).

Los modelos estocásticos, se basan en la probabilidad de sucesos discretos, los cuales pueden ser representados gráficamente y matemáticamente en redes de Petri y solucionarse mediante algoritmos como los establecidos por Gillespie que involucran el método de Montecarlo para su resolución (1,6). Sin embargo, las simulaciones estocásticas son computacionalmente más costosas. Para un sistema bien mezclado de volumen infinito (número de partículas infinitas), las fluctuaciones tienden a cero y el proceso aleatorio es bien aproximado por un modelo macroscópico determinista, mientras que, para sistemas no lineales, las fluctuaciones aleatorias pueden ser importantes (7).

Los algoritmos para el desarrollo de modelos estocásticos propuestos por Gillespie, se dividen en dos métodos matemáticos con formulaciones diferentes pero equivalentes, conocidos como el método de la primera reacción y el método directo, que se enmarcan en una función de densidad que determina la probabilidad de ocurrencia de un evento o reacción (8).

Las aproximaciones de Gillespie se pueden resumir en los siguientes pasos:

1. *Inicialización*: Inicializa el número de moléculas en el sistema, reacciones, constantes y generación de números aleatorios.
2. *Pasos de Monte Carlo*: Genera números aleatorios para determinar cuál es la siguiente reacción que ocurre, así

como el intervalo de tiempo. La probabilidad de escoger una reacción dada es proporcional al número de moléculas de sustrato.

3. *Actualización*: Incrementa el tiempo en un paso por la generación aleatoria de tiempo en el paso 2. Actualiza el conteo de moléculas basado en las reacciones ocurridas.
4. *Iteración*: Regresa al paso 2 a menos que el número de reactivantes sea cero o el tiempo de simulación se haya excedido.

En detalle el método se describe a continuación en la ecuación a:

$$(a) \quad P(\tau, \mu) = h_{\mu} * c_{\mu} \exp(-\sum_{\mu=1}^M h_{\mu} * c_{\mu} * \tau)$$

Donde $P(\tau, \mu)$ es la probabilidad en un tiempo t de que la siguiente reacción ocurra en un intervalo de tiempo $(t + \tau, t + \tau + d)$ y que esta reacción sea R_{μ} , es el tiempo estocástico. c_{μ} es una constante estocástica de reacción para cada reacción R_{μ} , que al ser multiplicado por h_{μ} que es el número de posibles combinaciones de R_{μ} genera la probabilidad a_{μ} de que R_{μ} se lleve a cabo, así:

$$(b) \quad a_{\mu} = h_{\mu} * c_{\mu}$$

reemplazando la ecuación (b) en la ecuación (a):

$$(c) \quad P(\tau, \mu) = a_{\mu} \exp(-\sum_{\mu=1}^M a_{\mu} * \tau)$$

El método directo y el de la primera reacción, se diferencian principalmente en la forma de determinar la probabilidad $P(\tau, \mu)$; ambos métodos son rigurosos y exactos, sin embargo, cuando el número de reacciones M es mayor a tres, el método directo es un poco más eficiente (8).

$$(d) \quad P(\tau, \mu) = P_1(\tau) * P_2(\mu/\tau)$$

De la ecuación principal que describe del método directo (ecuación d) se puede deducir que éste método genera en primera instancia un valor aleatorio τ (primer número aleatorio r_1) según $P_1(\tau)$ y posteriormente un número entero aleatorio μ (segundo número aleatorio r_2) según $P_2(\mu/\tau)$; el tiempo estocástico τ se calcula de acuerdo a la ecuación (e) y el segundo número aleatorio de distribución uniforme r_2 , determina cual reacción se llevará a cabo de acuerdo a la ecuación (f) (8):

$$(e) \quad \tau = \ln\left(\frac{1}{r_1}\right) * \left(\frac{1}{a_{\mu}}\right)$$

$$(f) \quad a_{\mu} * r_2$$

Por otro lado, el método de la primera reacción calcula el tiempo estocástico τ de acuerdo a la ecuación (g), donde r_1 es un número aleatorio de distribución uniforme. Los algoritmos completos para ambos métodos se describen con mayor exactitud en la sección de Materiales y Métodos.

$$(g) \quad \tau = \ln\left(\frac{1}{r_1}\right) * \left(\frac{1}{a_{\mu}}\right)$$

Las fluctuaciones aleatorias en la transcripción y traducción pueden llevar a estallidos temporales en el nivel de proteínas, asociado con una distribución no Poisson del número de moléculas y característica de correlaciones temporales. En general, las fluctuaciones aleatorias en sistemas vivos consisten de dos contribuciones: variación extrínseca debido a influencias por otros sistemas y variación intrínseca derivados de la dinámica aleatoria en el propio sistema (7).

Este trabajo se centra en evaluar un modelo de una red génica simple para la síntesis de una proteína, analizando la red de Petri producto del trabajo de Goss y Peccoud en 1998 (Figura 1, (1)). El modelo se analizó desde el punto de vista estocástico por el método de la primera reacción y el método directo de Gillespie, adicionalmente se analizó desde el punto de vista determinístico y se comparó el comportamiento dinámico con cada una de estas aproximaciones.

Materiales y Métodos

El planteamiento tanto del modelo biológico, como de los modelos matemáticos (determinístico y estocástico), se realizó de acuerdo con la red de Petri propuesta por Goss y Peccoud en 1998 (1), plasmada en su trabajo “Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets”.

Modelo Biológico

En la Figura 1 se esquematiza por medio de una red de Petri el modelo génico simple para la síntesis de una proteína.

En este modelo se contempla un gen inactivo (P_1) que por medio de una tasa de activación (λ) cambia el gen en estado inactivo a su estado activo (P_2); el gen activo (P_2) puede volver a su estado inactivo contemplando su tasa de inactivación (μ), estos dos procesos de activación-inactivación del gen son representados por las reacciones de las ecuaciones 1 y 2 respectivamente.

Reacciones del modelo donde $P_1=Gen\ inactivo$, $P_2=Gen\ activo$, $P_3=Proteína$

1. $P_1 \rightarrow P_2$
2. $P_2 \xrightarrow{\lambda} P_3$
3. $P_2 \xrightarrow{\mu} P_1$
4. $P_3 \xrightarrow{\sigma} Degra$

Inactive Gene

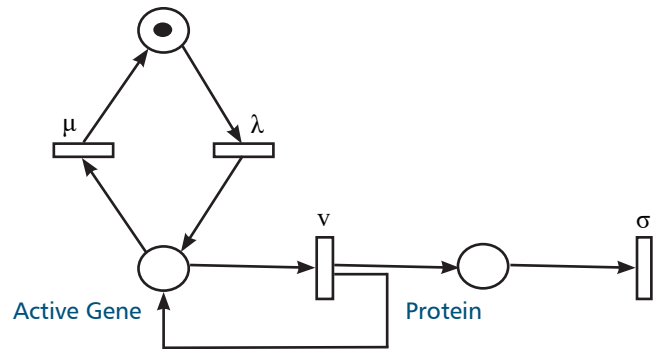


Figura 1. Representación de un modelo génico simple, tipo Red de Petri para la síntesis de una proteína (Goss y Peccoud, 1998).

El gen activo (P_2) sintetiza la proteína (P_3) con una tasa de síntesis (v). El gen activo se considera simultáneamente tanto como una entrada y como una salida para la síntesis (v), ya que se requiere para la síntesis, pero no es consumido, actuando efectivamente como un catalizador. Para la proteína P_3 se asume una tasa de degradación (σ), representada por la reacción de la ecuación 4. Los valores para los parámetros que se asumen en este modelo se muestran en la Tabla 1.

Tabla 1. Constantes de velocidad para el modelo génico simple para la síntesis de una proteína.

Descripción	Parámetro	Constante
Tasa de Activación del Gen (P_1)	λ	$1.0\ s_1$
Tasa de Inactivación del Gen (P_2)	μ	$5.0\ s_1$
Tasa de Síntesis de la Proteína (P_3)	v	$1.0\ s_1$
Tasa de degradación de la Proteína	δ	$0.02\ s_1$

Modelo Determinístico

Tomando como base el modelo biológico de la red génica simple para la síntesis de una proteína (Figura 1), las tasas de los parámetros correspondientes a cada una de las reacciones implicadas en el modelo (Tabla 1) y las condiciones iniciales (número de moléculas) para cada especie de la red (Tabla 2), se planteó el sistema de ecuaciones diferenciales que representan la interacciones entre las especies (Ver ecuaciones 5 a 7).

$$5. \quad \frac{dP_1}{dt} = -\lambda P_1 + \mu P_2$$

$$6. \quad \frac{dP_2}{dt} = \lambda P_1 - \mu P_2$$

$$7. \quad \frac{dP_3}{dt} = v P_2 - \sigma P_3$$

Para este modelo se asumió una cinética relativa a las ecuaciones de velocidad para las tres reacciones anteriores, que describen el sistema bajo los parámetros que asume la ley de acción de masas para todas las especies de la red en equilibrio dinámico. El balance de materia que considera la diferencia entre las reacciones de generación o producción de la especie y las reacciones de consumo de esta en la red fue considerado para todas las especies.

Tabla 2. Condiciones iniciales de las especies del modelo génico simple para la síntesis de una proteína.

Descripción	Especie (componente)	Concentración Inicial (moléculas)
Gen Inactivo	P ₁	1
Gen Activo	P ₂	0
Proteína	P ₃	0

Posterior al planteamiento matemático del sistema, fueron obtenidos los perfiles de concentración mediante la función ode45 para solución numérica de ecuaciones diferenciales de cada una de las especies del modelo, implementada en Matlab® (Versión R-2011a). Se varió el tiempo de ejecución con $t=10$ y $t=100$. Posteriormente los valores de las constantes fueron variados para tener una medida de la dependencia del sistema a diferentes variables.

Modelo Estocástico

Método de la Primera Reacción

Para la implementación de este método se siguió el siguiente algoritmo:

1. Determinación de las condiciones iniciales de los parámetros (Tabla 2) y de las Constantes cinéticas (Tabla 1) en $t=0$.
2. Cálculo de a_{μ} (ecuación b) para cada uno de los eventos (reacciones R_{μ}) que depende del tipo de reacción. Para este modelo se asumen en su totalidad reacciones monomoleculares por lo tanto se asume el valor de c_{μ} igual al de la constante cinética determinada para cada reacción.
3. Cálculo de τ (tiempo estocástico) para cada una de las reacciones. Para esto se determina un número aleatorio de distribución uniforme r_1 .
4. Selección del evento (Reacción) de menor tiempo.
5. Ejecución de la reacción de menor tiempo
6. Actualización del tiempo τ (el tiempo es incrementado en τ)

7. Actualización del inventario de moléculas.
8. Repetición de los pasos 1 a 7 hasta alcanzar el tiempo de simulación final ($\tau_{\text{Acumulado}} > \tau_{\text{Final}}$)

El algoritmo anterior fue planteado y resuelto en Matlab® (Versión R-2011a).

Método Directo

Para la implementación de este método se siguió el siguiente algoritmo:

1. Cálculo de a_{μ} para cada uno de los eventos (reacciones R_{μ}), teniendo en cuenta las condiciones iniciales de los parámetros (Tabla 2) y las constantes cinéticas del modelo (Tabla 1)
2. Calculo del a_T ($\sum a_{\mu}$ para cada una de las reacciones)
3. Se determina un número aleatorio de distribución uniforme r_1 y se calcula el tiempo τ (tiempo estocástico) para cada una de las reacciones (R_{μ}) mediante la siguiente expresión: $\tau = \frac{1}{a_T} \ln\left(\frac{1}{r_1}\right)$
4. Se genera un número aleatorio de distribución uniforme r_2 que esté entre $[0,1]$
5. Divido cada uno de los a_{μ} entre a_T y con r_2 evalúo en qué evento estoy. $\frac{a_{\mu-1}}{a_T} < r_2 \leq \frac{a_{\mu}}{a_T}$
6. Se ejecuta con el tiempo conocido τ y el evento conocido (calculado en el paso 5)
7. Se realiza inventario
8. Calcular el tiempo acumulado
9. Repetición de todos los pasos, a partir del paso 1 hasta que $\tau_{\text{Acumulado}} > \tau_{\text{Definido}}$

El algoritmo anterior fue planteado y resuelto en Matlab® (Versión R-2011a).

Análisis de Sensibilidad

Para llevar a cabo el análisis de sensibilidad, se variaron los valores de las constantes cinéticas un valor por encima y uno por debajo de los valores iniciales que se muestran en la Tabla 1. Este análisis se llevó a cabo para los tres métodos evaluados en este trabajo. Los valores empleados en el análisis de sensibilidad para la aproximación determinística y las dos aproximaciones estocásticas se especifican en la Tabla 7.

Para determinar si el efecto del parámetro sobre la media fue significativo, se realizó un análisis de varianza (ANOVA) empleando el software RStudio versión 3.1.0 de 2014. Se evaluó la

significancia estadística “p” bajo la hipótesis nula H_0 , con el fin de encontrar una diferencia entre los valores de la media de la proteína al variar las constantes cinéticas. Se utilizaron los 1000 valores medios en $t=100$ segundos obtenidos de los ciclos de ejecución del algoritmo para cada uno de los métodos estocásticos evaluados en este trabajo.

Resultados y Discusión

Cinética Determinística

Los resultados obtenidos a partir de la aproximación determinística del modelo, la cual está basada en la ley de acción de masas y que describe el sistema biológico con una gran cantidad de moléculas, donde las condiciones iniciales fijan el comportamiento del sistema se pueden evidenciar en la Figura 2. Al ser ejecutada la simulación a un tiempo $t=10$ segundos (figura 2-A) el aumento de la concentración de proteína (línea rosa) empieza a incrementar proporcionalmente con respecto al consumo del gen inactivo (línea azul) y por ende a la aparición del gen activo (línea morada), el cual iniciará el proceso de transcripción y traducción. A medida que se aumenta el tiempo de ejecución hasta $t=100$ segundos (figura 2-B), el comportamiento exponencial de la concentración de proteína se evidencia claramente. Los resultados arrojados son congruentes con el proceso biológico que enmarca la síntesis de proteínas.

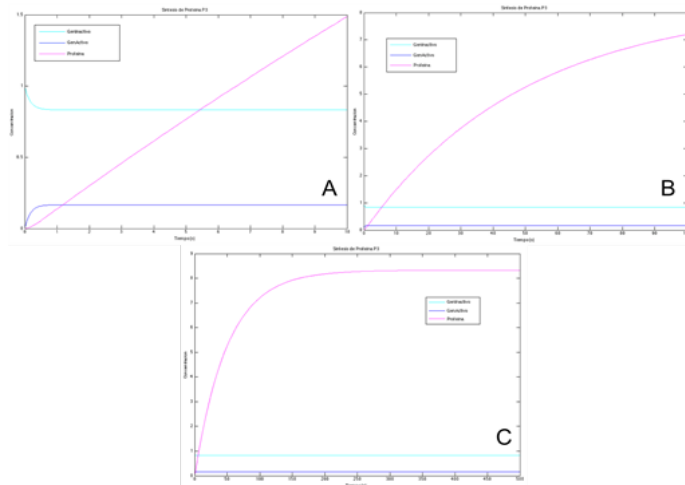


Figura 2. Modelo Determinístico de la red génica simple para la síntesis de una proteína. Perfil de concentraciones para P1, P2 y P3 a un tiempo A) $t = 10$ s en 0.919770s de simulación y B) $t = 100$ s en 1.017851 s. de simulación. C) $t = 500$ s en 1.88 s. La línea azul representa el gen inactivo, la línea morada el gen activo y la línea rosa representa la proteína. En el eje X se representa el tiempo en segundos y en el eje Y la concentración en número de moléculas.

En la figura 2C se evidencia como el sistema alcanza el estado estacionario a un tiempo aproximado de $t=200$ segundos, tardándose en resolver 1.88 segundos y donde la proteína alcanza un valor máximo de concentración de 8.3330 moléculas.

Vale aclarar que en la aproximación determinista de este modelo se ignora la variabilidad del sistema y la heterogeneidad que se presenta al analizar varias poblaciones celulares, razón por

la cual la simulación representa el comportamiento promedio de una sola célula. El modelo determinístico entonces, describe la variación neta de las especies (genes, mRNAs, proteínas) en el tiempo mediante un sistema de ecuaciones diferenciales. Éste modelo es una aproximación global que suele emplearse en sistemas con gran cantidad (altas concentraciones) de moléculas, en los que detallar cada uno de los eventos celulares que se llevan a cabo dentro de un organismo resulta en una alta demanda computacional, es así, como se considera que éste tipo de modelos son de carácter macroscópico, en los cuales no se detallan todos los eventos celulares característicos que ocurren bajo condiciones naturales (9). En general los resultados obtenidos por medio de esta aproximación determinística muestran el comportamiento global del modelo génico simple para la síntesis de una proteína.

Cinética Estocástica

El método de la primera reacción y el método directo, son dos métodos utilizados para representar una cinética estocástica, la cual se caracteriza por describir sistemas biológicos con baja concentración de especies, en donde las condiciones iniciales no fijan el comportamiento del sistema; es decir, para cada simulación que se realiza se obtiene un comportamiento diferente al ser una solución de la ecuación maestra en la que se utiliza el método de Montecarlo, donde cualquier número puede tener la probabilidad de salir en la generación de números aleatorios, además de ser procesos sin memoria. A continuación, se presentan los resultados y análisis de estos obtenidos con cada una de las aproximaciones evaluadas.

Método de la Primera Reacción

El modelo fue evaluado a 10, 100, y 1000 ciclos de ejecución del algoritmo planteado anteriormente para el método de la primera reacción y a 10 y 100 unidades de tiempo.

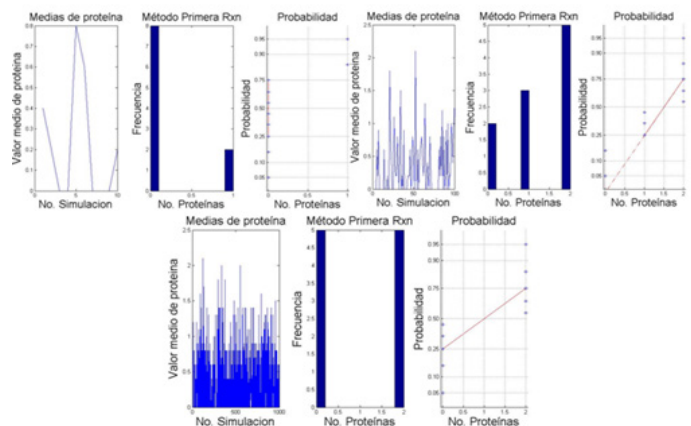


Figura 3. Valor Medio de la Proteína, Frecuencia y Probabilidad del Modelo Estocástico por el método de la Primera Reacción de la red génica simple para la síntesis de una proteína en tiempo $t=10$ segundos variando el número de ejecuciones del algoritmo de 10 (panel superior izquierdo), 100 (panel superior derecho) y 1000 (panel inferior).

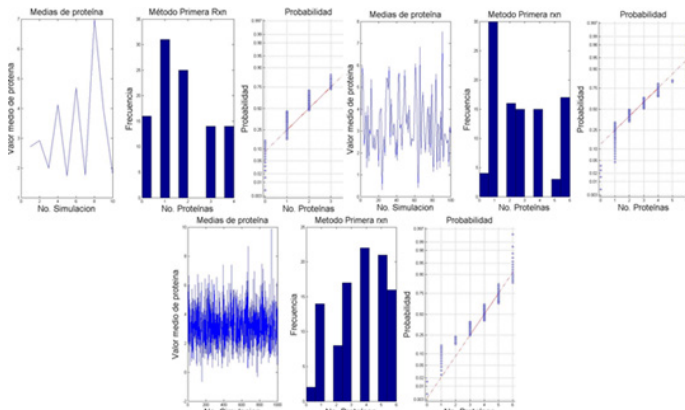
Fueron construidos los histogramas que reflejan los resultados obtenidos para la concentración de proteína en cada ciclo de ejecución (Figura 3 y 4). Los resultados de tiempo de simulación, media y desviación estándar obtenidos a partir de los datos de cada ciclo de ejecución se muestran en las tablas 3 y 4.

Al realizar los 10, 100 y 1000 ciclos de ejecución del algoritmo a un tiempo $t=10$ segundos (Figura 3) se observa claramente que el tiempo del ensayo no es suficiente para que la distribución del comportamiento del sistema se estabilice. Lo que se evidencia adicionalmente en el hecho que en la repetición de cada ciclo de ejecución las frecuencias y las probabilidades varían considerablemente. También cabe notar que el número medio de proteínas para los tiempos de 10 segundos y 100 segundos, difieren notablemente puesto que a medida que aumenta el tiempo de ensayo se tiene mayor probabilidad de aumento de la concentración de la proteína si no se ha llegado al estado estacionario, mostrando de esta manera que a medida que aumenta el número de ciclos de ejecución del algoritmo (emulando un aumento en la población celular) se genera un incremento en la estocasticidad del sistema. Por lo cual se concluye que la probabilidad de encontrar la proteína P3 en el tiempo $t=10$ segundos es muy baja (Tabla 3).

Tabla 3. Último valor de la Media, Desviación y Varianza para la producción de proteína en $t=10$ segundos a diferentes ciclos de ejecución del algoritmo asociado a la figura 3.

Ciclos de ejecución del algoritmo	Tiempo de Simulación	Media	Desviación
10	0.18771 s	0.2	0.4216
100	0.193840 s	1.3	0.8233
1000	1.488001 s	1.0	1.0541

Figura 4. Valor Medio de la Proteína, Frecuencia y Probabilidad del Modelo Estocástico por el método de la Primera Reacción de la red génica simple para la síntesis de una proteína en tiempo $t=100$ segundos variando el número de ciclos de ejecución del algoritmo de 10 (panel superior izquierdo), 100 (panel superior derecho) y 1000 (panel inferior).



Al incrementar el tiempo de simulación a $t=100$ segundos y llevar a cabo el mismo número de ciclos de ejecución del algoritmo planteado para el método de la primera reacción (Figura 4), la concentración de proteína reflejada en los histogramas muestra un comportamiento característico que emula el comportamiento propio de los sistemas biológicos, lo cual también se evidencia en las gráficas de probabilidad. Básicamente podemos afirmar que aun mayor tiempo de simulación y un mayor número de ciclos de ejecución del algoritmo se obtiene una mejor descripción del comportamiento del sistema.

Tabla 4. Último valor de la Media, Desviación para la producción de proteína en $t=100$ segundos a diferentes ciclos de ejecución del algoritmo, asociado a la figura 4.

Ciclos de ejecución del algoritmo	Tiempo de Simulación	Media	Desviación
10	0.039207 s	4,7	2,0913
100	0.242058 s	6,85	3,3586
1000	10.229929 s	3,39	1,786

De acuerdo con el análisis de los resultados que relacionan el valor medio de la proteína con el número de simulación (figuras 3 y 4) obtenidos empleando el método de la primera reacción y comparándolos con los resultados del método determinístico, se puede confirmar que la descripción del sistema mediante un sistema de ecuaciones diferenciales (método determinista) no representaría adecuadamente el comportamiento de la red, mientras que los resultados arrojados por el método estocástico de la primera reacción muestran en general una distribución de probabilidad que se acerca más al proceso biológico presente en la naturaleza.

Lo anterior está relacionado con una característica importante de los modelos de este tipo, la cual supone que el sistema de ecuaciones es continuo y determinístico, lo cual no se cumple en ningún fenómeno biológico, si se considera que la estocasticidad inherente en la naturaleza puede producir efectos no determinísticos importantes. En contraste, cuando el enfoque estocástico genera números aleatorios está considerando el comportamiento de cada molécula individual y a su vez las interacciones discretas que ocurren entre estas, por lo cual los modelos de Gillespie se ajustan mejor a la realidad en este y en la mayoría de los sistemas biológicos, confirmando nuevamente que nuestro modelo determinístico consigue predecir la cantidad final de proteína pero no es capaz de indicar los cambios de tendencia en el tiempo, lo cual está apoyado en el trabajo de Pahle en 2009 (10).

Método Directo

Bajo esta aproximación el modelo también fue evaluado a 10, 100, y 1000 ciclos de ejecución del algoritmo planteado anteriormente para el método directo y a 10 y 100 unidades de tiempo. Los histogramas que reflejan los resultados obtenidos para la

concentración de la proteína fueron construidos para cada ejecución del algoritmo (Figura 5 y 6). Los resultados de tiempo de simulación, media y desviación estándar obtenidos a partir de los datos de cada ciclo de ejecución del algoritmo se muestran en las tablas 5 y 6.

Tabla 5. Valores asociados a la figura 5. para el último valor de la Media, Desviación y Varianza de la producción de proteína en $t=10$ segundos a diferentes ciclos de ejecución del algoritmo.

Ciclos de ejecución del algoritmo	Tiempo de Simulación	Media	Desviación
10	0.052608	0.9	1.1005
100	0.406881	0.56	0.8204
1000	4.2368	0.518	0.7472

Para las simulaciones de 100 y 1000 iteraciones con ciclos de ejecución del algoritmo en un tiempo $t=100$ segundos cada una, (panel intermedio e inferior Figura 6) se puede observar que se logra reproducir el fenómeno de regulación biológica, pues los histogramas adquieren una forma de distribución característica presente en la mayoría de los fenómenos biológicos de expresión génica presentes en la naturaleza.

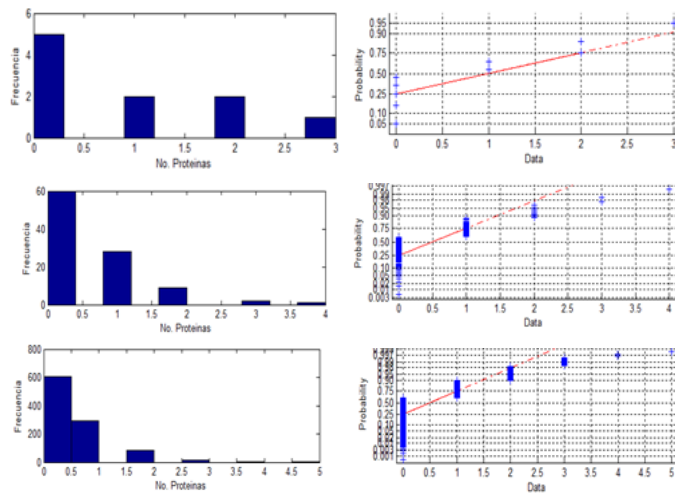


Figura 5. Frecuencia y Probabilidad calculadas a partir del método directo para la red génica simple de la síntesis de una proteína en tiempo $t=10$ segundos variando el número de ciclos de ejecución del algoritmo de 10 (Panel superior), 100 (Panel intermedio) y 1000 (Panel inferior).

Tabla 6. Valores asociados a la figura 6. para el último valor de la Media, Desviación y Varianza de la producción de proteína en $t=100$ segundos a diferentes ciclos de ejecución del algoritmo.

Ciclos de ejecución del algoritmo	Tiempo de Simulación	Media	Desviación
10	0.052608	0.9	1.1005
100	0.406881	0.56	0.8204
1000	4.2368	0.518	0.7472

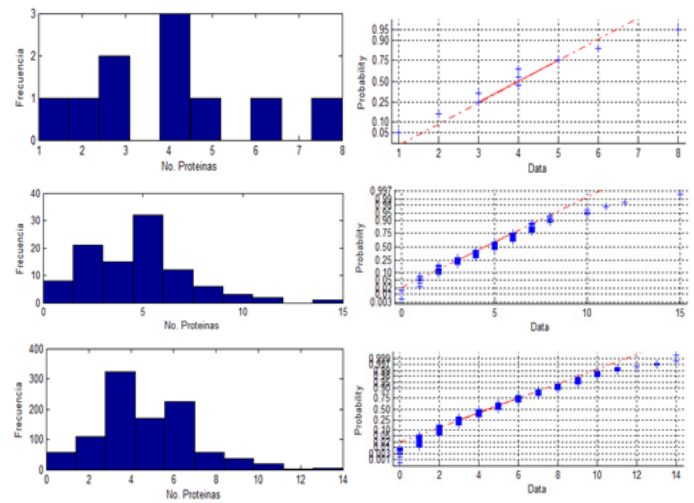


Figura 6. Frecuencia y Probabilidad calculadas a partir del método directo para la red génica simple de la síntesis de una proteína en tiempo $t=100$ segundos variando el número de ciclos de ejecución del algoritmo de 10 (Panel superior), 100 (Panel intermedio) y 1000 ciclos (Panel inferior).

La representación estocástica de las interacciones moleculares tiene una base teórica más fuerte que la cinética determinística cuando el número de moléculas es menor. Los resultados de un modelo estocástico difieren con su equivalente determinista, principalmente cuando se trata de la expresión génica, siendo necesaria la información cuantitativa para entender el fenómeno (1). Lo anterior posiblemente se debe al ruido transcripcional y traduccional propio de los sistemas biológicos y el cual es depreciado en los modelos deterministas.

De acuerdo con los resultados obtenidos en este trabajo, el método Determinístico es el que menor tiempo tarda en resolverse y el que requiere una menor demanda computacional; sin embargo, el método determinista no refleja el comportamiento de la población. Al comparar los resultados obtenidos a partir de las simulaciones de 1000 ciclos de ejecución del algoritmo y tiempo $t=100$ segundos para los dos métodos estocásticos evaluados, se tiene que el método de la primera reacción tarda ≈ 10.23 segundos, el método directo tarda ≈ 42.58 segundos para realizar la ejecución, en contraste el método Determinístico para $t=500$ segundos tarda 1.88 segundos.

Según las estructuras de los algoritmos lo esperado sería que el método directo tardara menos en realizar todas las iteraciones, pues en el método directo se generan únicamente dos números aleatorios, comparado con el método de la primera reacción en el que se generan 4 números aleatorios (uno para cada evento o reacción). Sin embargo, los resultados de las simulaciones muestran que el método directo es computacionalmente más demandante.

Lo que se cree fue causado por la manera en la que se planteó el algoritmo del método directo, pues en lugar de plantear un loop de tipo while se planteó un loop de tipo for encadenado con un condicional en el que el contador del ciclo se multiplica por un factor de 300, para asegurar que el ciclo termine solo cuando se dejara de cumplir la condición del comando interior.

Adicionalmente, al comparar los resultados representados en los histogramas que se obtuvieron por el método de la primera reacción (Figura 4) con los que se obtuvieron por el método directo (Figura 6), se observa que para métodos directos se requiere de un menor número de ciclos de ejecución del algoritmo para obtener un resultado confiable. Esto se evidencia en los histogramas del método directo (Figura 6) donde se observa que con 1000 ciclos de ejecución del algoritmo el resultado no varía significativamente con el observado a 100 ciclos de ejecución del algoritmo para el sistema que representa la expresión génica, lo cual no ocurre para el modelamiento estocástico cuando el sistema es simulado bajo la aproximación del método de la primera reacción (Figura 4) donde es necesario de tener el mayor número de ejecuciones del algoritmo para tener un resultado más confiable.

Cabe notar que el tiempo de simulación bajo la aproximación estocástica del sistema tanto para el método de la primera reacción como para el método directo es proporcional al número de eventos en el sistema. Se puede afirmar, por consiguiente, que estos algoritmos son de orden M , lo cual se representa que como $O(M)$; en otras palabras, si se duplica el número de eventos se esperaría que el tiempo de simulación se duplique de igual manera (11,12). En este punto es importante considerar que cuando el número de reacciones M es mayor a tres, el método directo es un poco más eficiente (8).

En adición, se puede inferir del planteamiento de los algoritmos presentados en la sección de materiales y métodos, que el método directo selecciona el evento que se efectúa mediante las desigualdades para a_i , en cambio el método de la primera reacción selecciona el evento mediante la identificación de τ_i más pequeño. Sin embargo, para los dos métodos se realiza una selección lineal (se verifica cada a_i y cada τ_i por medio de un condicional), lo que implica que se genera mayor gasto computacional a medida que se aumenta el número de reacciones, esto no representa un problema cuando se trata de un sistema con 4 eventos.

Pero cuando el número de reacciones sea mucho mayor, seguramente será más conveniente resolver el sistema de ecuaciones mediante el método directo si lo que se requiere es tener una aproximación más real a los fenómenos que ocurren a nivel biológico (Procesos transcripcionales y traduccionales) sin que la demanda computacional sea un limitante. Por otro lado, si lo que se requiere es observar el comportamiento global del sistema sin tener en cuenta los eventos biológicos de transcripción y traducción en un modelo de expresión génica lo recomendable es seguir un método determinístico que igualmente arrojará resultados confiables.

Análisis de Sensibilidad

Como se mencionó en el apartado de materiales y métodos se llevó a cabo el análisis de sensibilidad, se variando las constantes cinéticas un valor por encima y uno por debajo de los valores iniciales que se muestran en la Tabla 1. Lo anterior en busca de determinar el efecto que tienen los cambios en los valores de las constante cinéticas sobre la producción de la proteína para el

método determinístico y los dos métodos estocásticos evaluados en este trabajo. Los valores empleados en el análisis de sensibilidad para las tres aproximaciones mencionadas anteriormente se especifican en la Tabla 7.

Tabla 7. Valor de los parámetros utilizados

Parámetro	Constante
λ	0.8-1.2 s-1
μ	3.0-7.0 s-1
ν	0.7-1.3 s-1

El análisis de varianza (ANOVA) se realizó para determinar si el efecto del parámetro sobre la media fue significativo, los resultados arrojan un valor p bajo una hipótesis nula H_0 , lo cual indica que no existe una diferencia entre los valores de la media de la proteína al variar las constantes cinéticas, por lo que se rechaza si el valor de p es menor de 0.05.

Método Determinístico

Los resultados del análisis de sensibilidad siguiendo la aproximación determinística para el modelo se reflejan en la Figura 7 y la Tabla 8. Con estos resultados se evidencia que el cambio en las constantes cinéticas influye directamente en el número de proteínas total con mayor frecuencia. Cuando se aumenta el valor de λ el gen se activa más rápido reflejando una mayor producción de la proteína, mientras que si se aumenta el valor de μ la velocidad de la reacción reversible para tener el gen en su forma inactiva es mucho más grande que la constante cinética de la reacción hacia adelante (reacción que representa la transición del gen inactivo al gen activo), obteniendo así una menor producción de la proteína. Si se aumenta el valor de ν , es decir, la velocidad de producción de la proteína, ésta aumenta y sigue en aumento si la constante de degradación es menor.

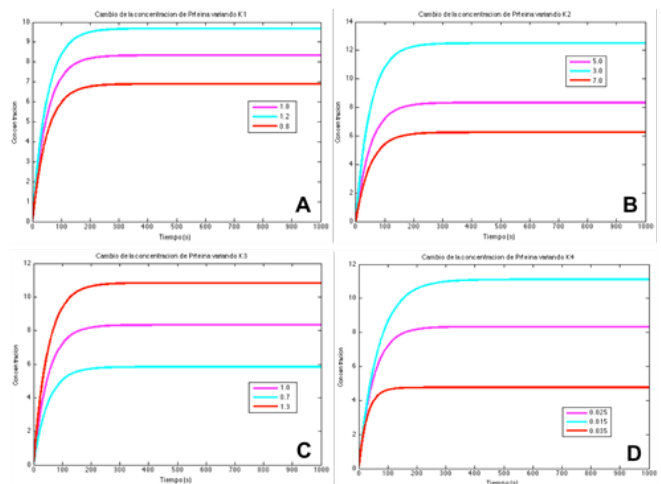


Figura 7. Modelo Determinístico de la red génica simple para la síntesis de una proteína. Perfil de concentraciones para P3 a un tiempo $t=1000$ s variando los parámetros A) variación de λ , B) Variación de μ C) Variación de ν y D) Variación de σ

El análisis de sensibilidad se realizó para las simulaciones de 1000 ciclos de ejecución del algoritmo y un tiempo $t=100$, debido a que estos ensayos son los que presentan menor dependencia al valor que tomara el número aleatorio. Sumado a esto las medidas estadísticas de media y desviación son más confiables debido al número de datos.

Tabla 8. Valores de la concentración de proteína en el modelo determinístico variando las constantes cinéticas en $t=500$ segundos. A condiciones normales la concentración de proteína fue 8.3330 moléculas.

	$\lambda=0.8$	$\lambda=1.2$
[Proteína]	6.8962	9.6770
	$\mu=3$	$\mu=7$
[Proteína]	12.4994	6.2497
	$v=0.7$	μ
[Proteína]	5.8331	10.8328
	$\sigma=0.015$	$\sigma=0.035$
[Proteína]	11.1050	4.7619

Método de la Primera Reacción

Los resultados del análisis de sensibilidad siguiendo la aproximación estocástica con el método de la primera reacción para el sistema biológico estudiado se reflejan en la Figura 8 y la Tabla 9.

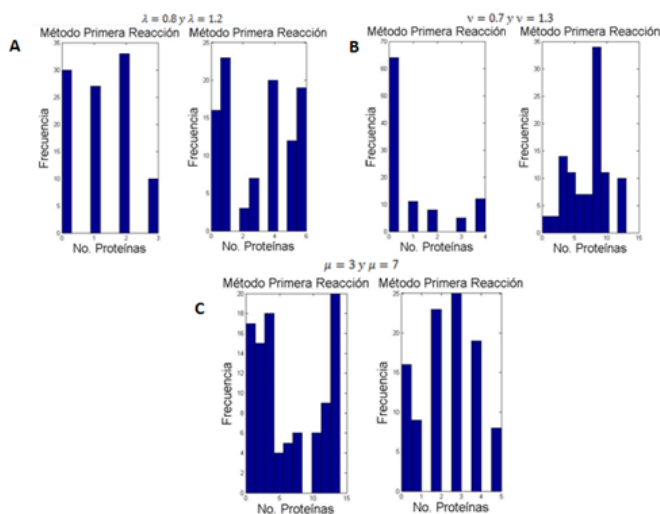


Figura 8. Histogramas para el Método de la Primera Reacción de la red génica simple para la síntesis de una proteína evaluado a tiempo $t=100$ y 1000 ciclos de ejecución del algoritmo variando los parámetros A) variación de λ , B) Variación de μ y C) Variación de ν

A partir de los resultados consignados en los histogramas de la Figura 8, se deduce que, al variar las constantes cinéticas, la frecuencia del número de proteínas varía (variación de la concentración). Para un valor mayor de λ hay mayor posibilidad de encontrar más cantidad de proteínas, pero con una frecuencia menor.

Por otro lado, al aumentar el valor de ν , el número de proteínas también es mayor, pero con una frecuencia más alta, lo cual biológicamente tiene sentido debido a que ν es la constante cinética de síntesis de la proteína. Al disminuir el valor μ la frecuencia para encontrar mayor número de proteínas es más alta debido a que existe mayor cantidad del gen en estado activo. Lo anterior se corrobora con los datos de los valores de p consignados en la Tabla 9.

Tabla 9. Resultados del Anova realizado en RStudio sobre la media a $t=100$ y 1000 ciclos de ejecución del algoritmo, variando parámetros usando el método de la primera reacción.

Parámetro	Media Cuadrada	Valor de F	Pr(>F)
λ	192.33	31.8	2.15e-14
μ	4887	742.1	<2.15e-16
ν	2178	365.2	<2.15e-16

Método directo

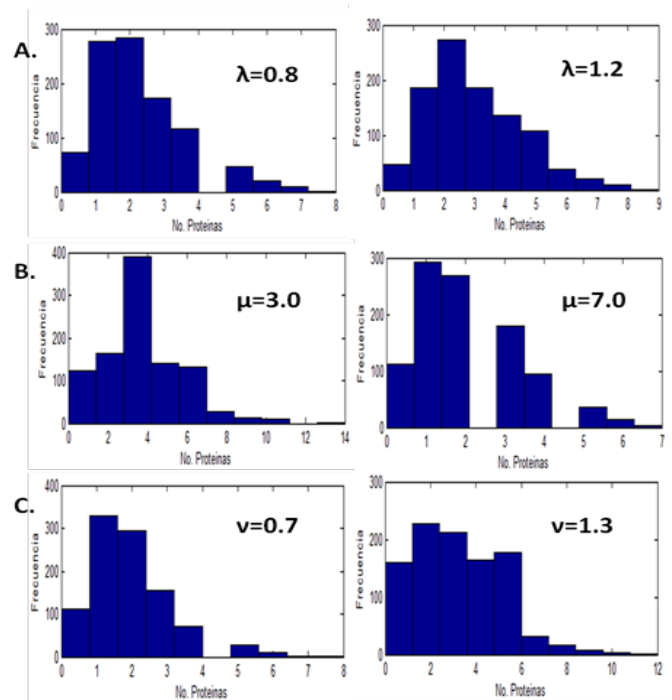


Figura 9. Histogramas para el Método directo de la red génica simple para la síntesis de una proteína evaluado a tiempo $t=100$ y 1000 variando ciclos de ejecución del algoritmo los parámetros A) variación de λ , B) Variación de μ y C) Variación de ν

Los resultados del análisis de sensibilidad siguiendo la aproximación estocástica con el método directo para el sistema biológico estudiado se pueden observar en la Figura 9 y la Tabla 10.

De acuerdo con los resultados mostrados en la tabla 9 y 10 se puede evidenciar que el valor de p para el método de la primera reacción y el método directo es menor a 0.05, por lo cual se rechaza la hipótesis, es decir, existe una diferencia significativa entre los valores medios de cantidad de proteína al variar las constantes cinéticas.

Tabla 10. Resultados del Anova realizado en RStudio sobre la media a $t=100$ y 1000 ciclos de ejecución del algoritmo, variando parámetros usando el método directo.

Parámetro	Media Cuadrada	Valor de F	Pr(>F)
λ	238.85	13.55	1.55e-6
μ	1362.1	50.06	<2e-16
ν	7351	268.7	<2.16e-16

Los resultados de las dos aproximaciones estocásticas no se ajustan en su totalidad a los datos reportados por Goss y Peccoud (1998) en su trabajo “Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets” (1), debido probablemente a que Goss y Peccoud realizan su trabajo aumentando el número de ciclos de ejecución de los algoritmos hasta un valor de 100000.

Conclusiones

El modelo determinista del sistema de ecuaciones diferenciales acopladas logra predecir los valores en estado estacionario acercándose a la aproximación reportada en el artículo de Goss y Peccoud de 1998 (1), pero no arroja información acerca de la distribución de los diferentes estados dentro de la población.

La resolución mediante el método directo propuesto por Gillespie fue la más demandante computacionalmente comparada con el método de la primera reacción y el método determinístico, lo que se cree que está relacionado con la manera en la que se planteó el algoritmo y por la generación de los dos números aleatorios que se requieren para la resolución del sistema.

Pese a lo anterior es el método directo el más recomendable para el estudio estocástico de modelos biológicos con un número de reacciones mayor a 3.

Mediante el análisis estadístico, se rechazaron todas las hipótesis nulas, es decir, se encontró que la cantidad de proteína total producida si depende de los valores que tomen las probabilidades de los eventos de activación del gen, inactivación del gen y producción de proteína.

Finalmente, se consiguió reflejar en este trabajo como los resultados obtenidos por medio de una aproximación determinística difieren de los arrojados al implementar una aproximación estocástica, esto principalmente justificado por el comportamiento global del modelo determinístico que tan solo refleja lo que ocurre a nivel de una sola célula, mientras que el estocástico considera el comportamiento de toda la población celular.

Referencias

- Goss PJ, Peccoud J. Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. *Proc Natl Acad Sci U S A* [Internet]. National Academy of Sciences; 1998 Jun 9; 95(12):6750–5. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9618484>
- Shmulevich I, Aitchison JD. Deterministic and Stochastic Models of Genetic Regulatory Networks. *Methods in enzymology* [Internet]. 2009. p. 335–56. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19897099>
- Mettetal JT, Muzzey D, Pedraza JM, Ozbudak EM, van Oudenaarden A. Predicting stochastic gene expression dynamics in single cells. *Proc Natl Acad Sci* [Internet]. 2006 May 9 103(19):7304–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16648266>
- Heiner M, Gilbert D, Donaldson R. *Petri Nets for Systems and Synthetic Biology. Formal Methods for Computational Systems Biology* [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. p. 215–64. Available from: http://link.springer.com/10.1007/978-3-540-68894-5_7
- Pedraza JM, van Oudenaarden A. Noise Propagation in Gene Networks. *Science (80-)* [Internet]. 2005 Mar 25;307(5717):1965–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15790857>
- Srivastava R, You L, Summers J, Yin J. Stochastic vs. deterministic modeling of intracellular viral kinetics. *J Theor Biol* [Internet]. 2002 Oct 7;218(3):309–21. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12381432>
- Klipp E (Edda), Liebermeister W, Wierling C, Kowald A. *Systems biology : a textbook*. 2nd ed. 2016. 488 p.
- Gillespie DT. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys* [Internet]. Academic Press; 1976 Dec 1;22(4):403–34. Available from: <https://www.sciencedirect.com/science/article/pii/0021999176900413>
- Machado D, Costa RS, Rocha M, Rocha I, Tidor B, Ferreira EC. Model transformation of metabolic networks using a Petri net based framework. Available from: https://www.kimosys.org/documentation/Machado_ModelReduction.pdf
- Pahle J. Biochemical simulations: stochastic, approximate stochastic and hybrid approaches. *Brief Bioinform* [Internet]. Oxford University Press; 2009 Jan 10(1):53–64. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19151097>
- Barnes DJ. *Introduction to modeling for biosciences*. Springer; 2014.
- Barnes DJ, Chu D. *Other Stochastic Methods and Prism. Introduction to Modeling for Biosciences* [Internet]. London: Springer London; 2010. p. 215–72. Available from: http://link.springer.com/10.1007/978-1-84996-326-8_6