



**CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS  
AVANZADOS DEL INSTITUTO POLITÉCNICO NACIONAL**

Unidad Monterrey

***Diseño in silico de circuitos genéticos a través de  
compuertas lógicas***

**Tesis que presenta**

Rubén Guzmán Ramírez

**Para obtener el grado de Maestro en Ciencias en:**

Ingeniería y Física Biomédicas

**Director de tesis**

Dr. Jesús Guadalupe Rodríguez González

Apodaca, Nuevo León

Agosto 2014





# ÍNDICE

---

	Página
Glosario	8
Resumen	13
Abstract	14
Introducción	15
Antecedentes	17
Biología sintética	17
Compuertas lógicas biológicas	19
Serín integrasas	21
Objetivo general	23
Objetivos específicos	23
Metodología	24
Sistema biológico	24
Modelo de los promotores inducibles	26
Modelo determinista	28
Modelo estocástico	32
Circuitos extendidos	34
Análisis de simulación	34
Resultados y discusión	36
Promotores inducibles	36
Análisis determinista	38

Análisis estocástico	43
Análisis de los circuitos extendidos	51
Factibilidad en la implementación experimental de los circuitos	59
Conclusiones	63
Apéndice A. Descripción del planteamiento estocástico	65
Referencias	67

# ÍNDICE DE FIGURAS

---

Figura		Página
1	Uso del transcriptor para construir compuertas lógicas con tres terminales.	25
2	Estados de unión posibles para los promotores PBAD y PTet.	27
3	Dinámica de los promotores PBAD y PTet.	37
4	Dinámica de la expresión de la integrasa A.	39
5	Reproducción de la tabla de verdad de la compuerta AND a través del modelo estocástico.	42
6	Simulación estocástica de la compuerta NAND.	44
7	Efecto del número de copias de los plásmidos portadores de la compuerta lógica AND.	46
8	Desempeño predictivo de los dispositivos lógicos.	49
9	Efecto del número de plásmidos en todas las compuertas.	50
10	Semi sumador.	52
11	Sumador completo.	53
12	Decodificador 2x4.	54
13	Expresión transitoria indeseada del sumador completo.	57

# ÍNDICE DE TABLAS

---

Tabla		Página
1	Parámetros para los modelos de los promotores PBAD y PTet.	28
2	Sistema de ODEs que describe la dinámica general de cada compuerta lógica.	29
3	Parámetros cinéticos usados en el modelo determinista.	30
4	Funciones lógicas para el control de la señal de salida.	31
5	Propensiones del algoritmo estocástico.	33
6	Ezimas pertenecientes a la familia de las serín integrasas.	61

# GLOSARIO

---

**Algoritmo estocástico:** Conjunto ordenado de operaciones que basa su resultado en probabilidades asociadas a sus variables que cambian en el tiempo.

**ARN polimerasa:** Se trata de una enzima de tipo ARN nucleotidiltransferasa. Su función es llevar a cabo la transcripción, proceso del que se obtiene una copia de ARN a partir de ADN según las leyes de complementariedad.

**ARNm:** Es el ácido ribonucleico que contiene la información genética procedente del ADN para la síntesis de proteínas. Es monocatenario, al contrario del ADN, que es bicatenario.

**Cinética enzimática:** Es el estudio de la velocidad de reacciones catalizadas enzimáticamente en función de la concentración de las moléculas involucradas, la presencia de inhibidores y algunos parámetros como la temperatura y el pH.

**Circuito:** En electrónica, se trata de un conjunto de elementos u operadores que, unidos entre sí, permiten establecer una corriente eléctrica entre dos puntos para aprovechar la energía eléctrica. En biología, el significado es similar aunque la señal y los componentes son de naturaleza bioquímica.

**Coefficiente de variación:** Es una medida de la dispersión relativa de un conjunto de datos, que se obtiene dividiendo la desviación estándar del conjunto entre su media aritmética y puede expresarse en términos porcentuales.

**Compuerta AND:** circuito de al menos dos entradas cuya operación lógica es un producto entre ambas, no es un producto aritmético, aunque en este caso coincidan. Su salida será alta si sus dos entradas están a nivel alto.



**Compuerta lógica:** dispositivo electrónico de conmutación que representa la expresión física de un operador booleano en particular.

**Compuerta NAND:** Responde a la inversión del producto lógico de sus entradas. Es decir, al igual que la función AND realiza la multiplicación de sus entradas, pero entrega el valor negado. En su representación simbólica se reemplaza la compuerta NOT por un círculo a la salida de la compuerta AND.

**Compuerta NOR:** La compuerta NOR realiza la función de suma, pero entrega el resultado invertido, ahorrando un NOT. Su salida será 1 solo si las dos entradas son 0.

**Compuerta NOT:** Dispositivo que invierte el valor que se le entrega. También tiene la utilidad de ajustar niveles pero tomando en cuenta que invierte la señal.

**Compuerta OR:** La compuerta OR realiza la función de suma lógica. Cuando se le aplica un uno a cualquiera de sus entradas el resultado de salida será uno, independiente del valor de la otra entrada. Excepto cuando las dos entradas estén en 0 la salida será 0.

**Compuerta XNOR:** Este dispositivo XNOR o Nor exclusivo, genera un resultado de salida que será alto si las dos entradas son del mismo valor, sean 0-0 ó 1-1. Es simplemente la inversión de la compuerta XOR.

**Compuerta XOR:** Esta compuerta XOR (or-exclusiva) se comporta de una manera especial. Su característica especial es que el resultado de salida será 1 si las dos entradas son distintas, sean 0-1 ó 1-0.

**Desviación estándar:** Es la raíz cuadrada de la media de los cuadrados de las puntuaciones de desviación de un conjunto de datos, es decir, la raíz cuadrada de la varianza de este conjunto.

**Digitalización:** Proceso de conversión de señales analógicas a señales digitales con el propósito de facilitar su procesamiento y hacer la señal resultante más resistente al ruido y otras interferencias.

**Dinámica:** Es el estudio que describe la evolución en el tiempo de un sistema en relación con las causas o condiciones que provocan cambios dentro del mismo.

**Dispositivo:** Es un aparato o mecanismo que desarrolla determinadas acciones. Su nombre está vinculado a que dicho artefacto está dispuesto para cumplir con su objetivo.

**Ecuación diferencial ordinaria:** Ecuación en la que aparecen derivadas ordinarias de una o más variables dependientes respecto a una única variable independiente.

**GFP:** Proteína monomérica de unos 230 aminoácidos responsable natural de las propiedades bioluminiscentes de la medusa *Aequorea victoria*. Puede usarse como una señal luminosa capaz de expresarse en células mediante las técnicas rutinarias de transgénesis.

**In silico:** Expresión utilizada para referirse a aquellos experimentos y procesos realizados por computadora o por simulación computacional.

**In vitro:** Condición artificial en la que se mantienen estructuras o componentes biológicos funcionales fuera del ambiente natural del cual fueron aislados.

**In vivo:** Condición de experimentación y observación fenomenológica que tiene lugar directamente sobre organismos vivos.

**Integrasa:** Es un tipo de enzima recombinasa obtenida naturalmente de bacteriófagos. Tiene la capacidad de catalizar la reacción de recombinación entre dos secuencias específicas de reconocimiento pertenecientes a moléculas separadas de ADN.

**Media:** Es el valor característico de una serie de datos que parte del principio de la esperanza matemática, se obtiene a partir de la suma de todos sus valores dividida entre el número de sumandos.

**Operador:** En el contexto de regulación de la expresión genética, es un elemento de control compuesto por una secuencia de ADN que es reconocida por proteínas reguladoras de la transcripción. Se sitúa entre la región promotora y los genes estructurales.

**Parámetro:** Es un factor o cantidad constante que caracteriza numéricamente a un objeto o a una variable dentro de un sistema.

**Plásmido:** Molécula circular de ADN que se replica de manera independiente al cromosoma de la célula hospedera. Su estructura se puede utilizar como vector de clonación para introducir ADN a las células de manera artificial.

**Promotor:** Es una región del ADN de características especiales que determina el punto en el que la ARN polimerasa comienza a transcribir un gen.

**Propensión:** Probabilidad asociada al acontecimiento de un evento. En el contexto de este trabajo, este evento consiste de una reacción bioquímica específica dentro una colección de ellas.

**Propiedad emergente:** Es una propiedad o atributo de una comunidad o sistema que no se aprecia en los individuos que las componen y que se hace evidente únicamente dentro del contexto poblacional en un espacio dado.

**Punto de equilibrio:** En un sistema dinámico, representa las condiciones de las variables para las cuales el sistema se encuentra estático. Para este trabajo, se evalúa la estabilidad de las variables a través del tiempo.

**Rango dinámico:** En sistemas digitales o mecanismos, es la proporción de máximos y mínimos niveles de señal entre los cuales se exige que mantengan una proporción de error especificada.

**Ruido biológico:** Proceso estocástico que produce variabilidad en la expresión de características de naturaleza biológica en sistemas virtualmente idénticos, como es el caso de la expresión genética.

**Señal de entrada:** Portador físico de información que será procesado por algún dispositivo específico.

**Señal de salida:** Portador físico que lleva la información después de haber sido procesada por algún dispositivo.

**Sistema:** Es un módulo ordenado de elementos que se encuentran interrelacionados y que interactúan entre sí. El concepto se utiliza tanto para definir a un conjunto de conceptos como a objetos reales dotados de organización.

**Terminador transcripcional:** Fragmento de ADN que determina el punto donde se disocia el complejo mRNA-RNA polimerasa-DNA, terminando por lo tanto la transcripción.

**Variable:** Término para designar una cantidad susceptible de tomar distintos valores numéricos dentro de un conjunto de números especificados.

# RESUMEN

---

En el presente trabajo se describe el diseño *in silico* de un conjunto de compuertas lógicas anteriormente formuladas de manera experimental bajo el nombre de transcriptor que permiten el control del flujo de la RNA polimerasa a través del ADN mediante la acción de serín integrasas. El objetivo principal de este estudio es la adquisición de información predictiva por medio de una aproximación teórica, facilitando así la ingeniería de dispositivos lógicos en un contexto biológico. Para ello se establecieron modelos deterministas y estocásticos que describen la dinámica del procesamiento de información de las compuertas. Adicionalmente se ensamblaron tres circuitos biológicos virtuales usando las compuertas lógicas como bloques de construcción: un semi sumador, un sumador completo y un decodificador 2x4. Se encontró que es posible obtener información predictiva a través de los modelos *in silico* tanto para las compuertas aisladas como para los circuitos. Ello permitió la anticipación de problemas relacionados a la naturaleza inherente de la acción de las integrasas y del ambiente interno de la célula que se manifiestan como expresión genética transitoria indeseable y que provocan la respuesta errónea de los dispositivos lógicos. Dichas predicciones permitieron proponer cambios en los circuitos implementados mediante la modificación de las partes biológicas y la alteración de la arquitectura de los dispositivos construidos para así resolver los dilemas encontrados.

Los resultados sugieren que las aproximaciones teóricas son útiles para la detección de posibles inconvenientes del ensamble experimental antes de llevar los circuitos a su construcción biológica.

# ABSTRACT

---

Here we depict an *in silico* design of a set of logic gates based on an already exposed three-terminal architecture named the transcriptor that uses bacteriophage serine integrases to control the flow of RNA polymerase along DNA. The main objective of this work is the acquisition of predictive data by means of a computer simulation approach facilitating the logic circuits engineering. It was established a mathematical model representing the dynamics of the logic gate's components using the deterministic and stochastic approximations. Three biologic virtual circuits were constructed: a half adder, a full adder and a 2x4 decoder. It was found that it is possible acquire predictive information from models before the circuit construction phase. This allows anticipation of issues related to biologic systems such as undesirable transitory genetic expression which leads to the wrong circuitry response. Such predictions allowed the proposal of solving strategies trough modification of the biologic parts and the devices architecture.

Results suggest that theoretical approximations are useful for the issues detection a priori circuits experimental assembly.

# INTRODUCCIÓN

---

El diseño y producción de elementos computacionales contruidos a partir de unidades biológicas con funcionamiento similar al de los dispositivos electrónicos es un esfuerzo conjunto de los campos de biología de sistemas y biología sintética (Barret *et al.*, 2006). Este último campo, desde su aparición, ha venido inspirando la visión de integrar el análisis computacional, información de índole biológica y el paradigma de ingeniería de sistemas en el diseño de nuevas “máquinas biológicas”. Dichas máquinas biológicas se originan de componentes biomoleculares básicos análogamente a como ocurre con los dispositivos eléctricos, y el flujo de información entre estos componentes requiere de un incremento en el conocimiento biológico junto con el poder de una aproximación apropiada de manejo de información.

En teoría, y desde la perspectiva de la biología sintética, el investigador debería ser capaz de iniciar con un conjunto de características deseadas en mente, diseñar la circuitería biológica que satisfaga esos requerimientos, e implementar el diseño *in vivo*. Sin embargo, la práctica real no es tan directa. La producción de sistemas biológicos complejos usualmente requiere optimización iterativa debido procesos celulares que pueden ser propios de los circuitos diseñados o ajenos a ellos. Adicionalmente, la circuitería sintética también sufre los efectos del ruido biológico y de condiciones iniciales indeseadas. Los problemas inherentes a este campo se vuelven más evidentes cuando se considera que los componentes biológicos, cuando interactúan entre sí, dan lugar a propiedades emergentes. La existencia de propiedades emergentes es un indicativo de que el conocimiento de los sistemas biológicos y las capacidades de diseño aún carecen del nivel de sofisticación requerido para generar diseños *a priori* y prototipos con las mejores probabilidades de éxito. Por ello, la construcción eficiente en biología sintética exige un medio para mejorar el diseño de cualquier dispositivo biológico con la capacidad de predecir propiedades y comportamientos anómalos a los propósitos para los cuales fueron creados.

Una manera de abordar a las estructuras biológicas desde la perspectiva de la ingeniería de sistemas consiste en la digitalización de la información. Para ello, normalmente se utiliza el sistema binario. La manipulación de la información binaria se realiza mediante circuitos electrónicos denominados compuertas lógicas, de las cuales hay una variedad y cada una de ellas formaliza una operación lógica diferente (Mano, 1994). De la forma en que en electrónica las compuertas lógicas permiten construir circuitos digitales, las redes de regulación genética codifican operaciones lógicas que son utilizadas por las células para integrar múltiples señales ambientales e internas (Lim, 2010) y responder de manera apropiada a ellas. Por lo tanto, mediante el diseño de circuitos genéticos que permitan vincular distintos sensores y actuadores celulares, es posible programar organismos unicelulares capaces de generar el comportamiento deseado en respuesta a señales específicas de entrada de naturaleza intra o extra celular (Khalil y Collins, 2010).

Existe una gran cantidad de reportes de compuertas lógicas biológicas funcionales en un contexto celular. Cada uno de ellos presenta diversas estrategias en que se involucran las áreas de la biología sintética y la biología de sistemas. El trabajo de Bonnet (Bonnet *et al.*, 2013) es un excelente ejemplo de ello y propone un sistema de procesamiento de información a través de la modificación específica al material genético en bacterias dependiendo de la combinación de señales de entrada y la operación lógica en cuestión. De manera experimental demostró que la arquitectura de seis compuertas lógicas funcionó de manera similar a un dispositivo electrónico. El paso subsecuente y obvio a esta idea es el uso de la colección de compuertas propuestas para el ensamblaje de redes más complejas, y con ello aumentar su versatilidad y aplicación como dispositivos de procesamiento y control celular. Sin embargo, y como se mencionó anteriormente, los circuitos biológicos de mayor complejidad frecuentemente muestran propiedades indeseadas o inesperadas. Para ello existe la posibilidad de evaluar su funcionamiento mediante modelos suficientemente realistas como para representarlos con herramientas informáticas previamente a su construcción *in vitro* o *in vivo*.

El propósito de este trabajo es la contribución de modelos *in silico* de estas construcciones artificiales y examinar su comportamiento y obtener predicciones funcionales a través de su implementación. Se propusieron así, tres circuitos construidos a partir de la colección de compuertas reportadas por Bonnet pero inspirados en la electrónica, y se evidenció la ventaja de generar modelos predictivos.



# ANTECEDENTES

---

El campo de la biología sintética es un campo emergente y apenas ha comenzado a desarrollarse, sin embargo existe una gran cantidad de trabajos al respecto cuyo objetivo principal es el de diseñar y fabricar nuevos componentes, redes y rutas biomoleculares, y utilizarlas para conectar y reprogramar organismos vivos. En este trabajo nos centramos en uno de ellos, mismo que describe una colección de compuertas lógicas cuyo cómputo es realizado a través de la acción de integrasas como herramientas moleculares. El objetivo es establecer un modelo *in silico* que permita reproducir los resultados que se han obtenido de manera experimental y posteriormente generar datos predictivos en el proceso de diseño de circuitos de mayor complejidad. A continuación se explicará brevemente el contexto de esta investigación en las áreas involucradas.

## ❖ Biología sintética

**Descripción:** Como se ha mencionado, la biología sintética combina los campos de la biología molecular y la ingeniería para generar sistemas biológicos artificiales. La conectividad de partes biológicas y su habilidad para procesar operaciones lógicas fue apreciada por primera vez hace más de 50 años (Monod y Jacob, 1961). Esta observación inspiró la descripción de esquemas de regulación biológica mediante modelos matemáticos (Glass y Kauffman, 1973; Savageau, 1974) y la aplicación de analogías de circuitos desde el marco teórico de la ingeniería eléctrica (McAdams y Arkins, 2000). Paralelamente, algunos logros en investigación genómica e ingeniería genética fueron edificando el inventario de los métodos necesarios para la construcción física de partes biomoleculares. Como resultado, la biología sintética nació con la perspectiva de generar la habilidad de conectar circuitos biológicos (ya sean de carácter genético, proteico, viral, genómico o como ruta de algún tipo) para la manifestación de formas lógicas de control celular. Hasta ahora, equipada con los conceptos ingenieriles de modularización, racionalización y modelaje, esta área ha progresado

de manera importante y ha generado ya toda una creciente batería de dispositivos genéticos y módulos biológicos. Si bien ello proporciona una gran oportunidad para evaluar el entendimiento que tenemos de los mecanismos biológicos naturales, también otorga la oportunidad de diseñar estructuras biológicas funcionales (Endy, 2005).

El primer paso en el proceso de ingeniería es la elección de una función biológica y la concepción de un sistema capaz de cumplir con esta función. Posteriormente se determinan los componentes que pueden ser usados para su construcción, y entonces se evalúa el desempeño del sistema (Fritz *et al.*, 2010). De tal manera que el objetivo es construir dispositivos biológicos personalizados que sean capaces de reconocer una o varias señales de entrada, procesar la información y realizar la función esperada. Los dispositivos con la posibilidad de comunicarse entre sí pueden ser colocados en el contexto celular apropiado y llegar así a la obtención de módulos biológicos y a la ejecución de tareas complejas.

El diseño exitoso y la construcción de las primeras redes genéticas sintéticas, el interruptor genético (Gardner, Cantor y Collins, 2000) y el *repressilator* (Elowitz y Leibler, 2000) demostraron que las metodologías inspiradas en ingeniería electrónica podían ser aplicadas en la obtención de sistemas biológicos dotados de comportamientos similares a los de un computador (memoria y cronometraje, respectivamente). Dentro del mismo marco teórico, se pueden obtener circuitos o redes a partir de partes biológicas bien definidas y estudiar su comportamiento mediante modelos predictores de la dinámica de un sistema (Cheng y Lu, 2012). Se ha aplicado en repetidas ocasiones el esquema anterior en la construcción de interruptores genéticos, elementos de memoria, osciladores generadores de pulsos, compuertas lógicas digitales, filtros y módulos de comunicación (Khalil y Collins, 2010).

**Limitaciones:** Aunque el control preciso en los sistemas biológicos es uno de los principales propósitos de la biología sintética existen diversas dificultades que surgen en la ingeniería dentro del contexto biológico y no así en los sistemas electrónicos (Hoffman-Sommer *et al.*, 2012). Una de ellas se desprende del hecho de que los mecanismos biológicos que realizan funciones deseables raramente se conocen a gran detalle. Adicionalmente, el ambiente tiene un gran impacto en la funcionalidad de los sistemas biológicos y, a nivel celular, cambios en el espacio extracelular pueden llevar a la pérdida de una función esperada. A pesar de estas dificultades muchos investigadores

han tenido éxito en la manipulación y extensión de rutas metabólicas y de señalización celular. Por ejemplo, se ha demostrado como usar algunas conexiones sintéticas de retroalimentación para reprogramar respuestas celulares (Bashor *et al.*, 2010).

**Aplicaciones:** El potencial de la biología sintética como campo ha sido demostrado en diferentes contextos. El rango de aplicación se extiende desde la terapéutica y el biosensado hasta la producción de biomateriales y biocombustibles. Por ejemplo, se sostiene el uso potencial de rutas sintéticas para la producción del precursor de fármacos para tratar afecciones como la malaria (Chen y Lu, 2010). Existen biosensores capaces de determinar la presencia de metabolitos y compuestos químicos o también señales físicas como la luz (Tabor *et al.*, 2009). También se han concebido circuitos destinados al tamizaje de compuestos como posibles candidatos de fármacos y construcciones genéticas que ofrecen control de expresión en mamíferos con fines de tratamiento terapéuticos (Stavreva *et al.*, 2009) o entrega controlada de fármacos (Weber *et al.*, 2007). En el campo de materiales hay propuestas interesantes, como la construcción de redes para la fabricación y recuperación de monómeros proteicos de fibras de seda de araña en sistemas procariontas (Widmaier *et al.*, 2009) o la organización de capas celulares de bacterias organizadas en patrones específicos para el desarrollo de estructuras a nivel macro o su uso en ingeniería de tejidos (Basu *et al.*, 2005). Incluso se han propuesto aplicaciones tan intrépidas como la ingeniería de bacterias competentes para el reconocimiento de tejido tumoral y la invasión de células cancerosas (Anderson *et al.*, 2006).

#### ❖ **Compuertas lógicas biológicas**

Las compuertas lógicas son dispositivos que implementan funciones booleanas y permiten convertir entradas lógicas en una sola salida. Para ello, las señales deben ser de naturaleza digital y generalmente se usa el sistema binario para cumplir este objetivo. En el sistema binario existen dos valores posibles: el primero, que es 0/falso, denota una señal baja o totalmente ausente; el segundo, que es 1/verdad, representa una señal alta. Aunque el concepto se refiere principalmente a señales de voltaje en redes electrónicas, también es posible fabricar compuertas lógicas biológicas de partes celulares como lo son los circuitos genéticos (Moon *et al.*, 2012).

Aunque existen sistemas biológicos naturales que procesan datos a través de operaciones lógicas (Morris *et al.*, 2010), también es cierto que algunas de ellas se encuentran representadas en mayor medida que otras, o bien, la construcción de dispositivos para algunas funciones es más sencilla que la formalización de otras. Como resultado, es más común encontrar en la literatura ensamblajes sintéticos que ejecutan las operaciones AND, OR y NOT (Tamsir, Tabor y Voigt, 2011). Una idea interesante en respuesta a lo anterior es la conexión entre dispositivos para generar funciones difíciles de construir. Por ejemplo, se dice que las compuertas NOR y NAND son únicas ya que son funcionalmente completas, lo que significa que cualquier operación computacional puede ser implementada conectando cualquiera de ellas entre sí (Scharle, 1965). Explotando esta posibilidad, se han realizado esfuerzos para crear una gama completa de compuertas lógicas biológicas a partir de unas pocas operaciones (Tamsir, Tabor y Voigt, 2011; Moon *et al.*, 2012). También existen diseños que permiten reconfigurar una sola construcción para generar múltiples funciones variando las señales de entrada sin tener que modificar físicamente el dispositivo (Goñi-Moreno y Amos, 2012).

Aunque hay múltiples esfuerzos de digitalización celular, actualmente es mucho más complicado construir compuertas biológicas que su contraparte en circuitos electrónicos. Una de las mayores dificultades se origina de la necesidad de conectar las unidades biológicas que ejecutan tareas específicas. En este sentido, la implementación de compuertas lógicas dentro de una célula se acopla con diversas limitaciones, como lo son la interferencia entre rutas de señales sintéticas y la propagación descontrolada del ruido. Algunos grupos han atacado dificultades de esta índole mediante algunas estrategias como lo es la introducción de compartimentalización celular. Por ejemplo, Regot y colaboradores (2011) propusieron e implementaron diferentes cepas de *Saccharomyces cerevisiae* donde cada una de ellas realiza una sola función lógica y responde a una señal de entrada particular. Estas tácticas permiten evitar la interferencia entre las señales de las compuertas y aminoran en cierta medida el ruido en etapas intermedias aunque también disminuyen la habilidad de conectar dispositivos entre sí y complica la construcción de circuitos más complejos.

Normalmente se encuentran compuertas que procesan información a partir de señales compuestas de activadores o factores de transcripción, o bien, a través de regulación a nivel proteico. Sin

embargo, un planteamiento interesante es el uso de integrasas para dicho fin. Bonnet y colaboradores (2013) propusieron un conjunto de seis compuertas lógicas codificadas bajo una misma arquitectura que denominaron transcriptor. El nombre deriva de su similitud con la estructura de un transistor, puesto que también consta de tres terminales: una para una señal de entrada, otra para una señal de salida y la última para dos señales de control constituida por dos integrasas diferentes.

El presente trabajo hace un estudio *in silico* utilizando la arquitectura del transcriptor ya que se presenta como una colección de funciones lógicas sin la necesidad de ensamblar compuertas en serie como se ha propuesto anteriormente, lo que disminuye su complejidad y facilita su construcción física. También se demostró un rango dinámico experimental similar para cada una de los dispositivos. Adicionalmente, las señales de entrada y salida son de la misma naturaleza por lo que es posible respaldar la conexión de compuertas dentro y entre dispositivos.

### ❖ Serín integrasas

Debido a que en este estudio interviene la representación virtual de la acción catalítica de las serín integrasas, a continuación se presenta una breve descripción de las mismas.

Las integrasas aisladas de bacteriófagos son enzimas que median la recombinación sitio-dirigida entre dos secuencias de reconocimiento en el ADN.

De entre la familia de recombinasas, las integrasas tienen la característica especial de que la direccionalidad en la reacción de recombinación es influenciada por un cofactor denominado excisionasa. En sistemas naturales, una integrasa típicamente cataliza la recombinación sitio-específica entre un sitio de acoplamiento en el genoma del fago infectante (attP) y un sitio de acoplamiento codificado dentro del genoma de la bacteria anfitriona (attB). La reacción de integración resultante inserta el genoma del fago dentro del cromosoma de la bacteria y al finalizar queda flanqueado por los sitios recién formados attL y attR (sitios LR). Después de un proceso de inducción para la fase lítica, el profago coexpresa la integrasa y excisionasa que, en conjunto, restauran la condición independiente del genoma del fago y los sitios originales attB y attP (sitios BP) (Ptashne, 2004).

Algunos trabajos con la mutación polar r32 del fago lambda revelaron que la recombinación mediada por fagos en sitios BP anti-paralelos (dispuestos en dirección contraria) lleva a la inversión del ADN involucrado (Reyes et al., 1979) en lugar de generar integración o supresión. Posteriormente se demostró, por experimentos *in vitro*, que si se expresa el complejo de integrasa y excisionasa conjuntamente se puede entonces revertir una secuencia de ADN flanqueada por sitios LR situados de manera opuesta (Pollok y Nash, 1983).

Las recombinasas de fagos están representadas en dos familias evolutivas distintas de la familia de recombinasas (Stoll, Ginsburg y Calos, 2002). Las tirosín integrasas, como la del bacteriófago lambda, generalmente presentan largos sitios de acoplamiento (~200 bp) y requieren cofactores específicos del anfitrión. Por otro lado, las serín integrasas pueden tener sitios de acoplamiento más cortos (~50 bp) y algunas de ellas no requieren ningún cofactor, lo que ha permitido usarlas exitosamente en un rango variado de organismos (Bonnet *et al.*, 2012). Debido a estas características se exploró la idea de la ingeniería de sistemas genéticos de almacenamiento de información re escribibles utilizando la familia de las serín integrasas.

# OBJETIVO GENERAL

---

- Realizar el diseño *in silico* de circuitos genéticos de múltiples entradas utilizando compuertas lógicas controladas por integrasas.

# OBJETIVOS ESPECÍFICOS

---

- Construir un modelo *in silico* para la descripción del comportamiento de las compuertas lógicas.
- Proponer un circuito genético de múltiples entradas y evaluar la factibilidad de su implementación.

# METODOLOGÍA

---

## ❖ Sistema biológico

Son seis compuertas lógicas las que se representaron a través del modelo *in silico*. Cada una de ellas comparte una misma arquitectura con una estructura interna que las diferencia. La arquitectura fue denominada transcriptor (Bonnet *et al.*, 2013) y cuenta con dos terminales para las señales de entrada y de salida, y una más para las señales de control que determinan el estado de cada compuerta lógica (Figura 1 panel A). Río arriba (*upstream*) de las compuertas se encuentra un promotor transcripcional constitutivo que permite el inicio del flujo de la RNA polimerasa y provee la señal de entrada. Por otro lado, río abajo (*downstream*) se sitúa el gen codificante para la proteína verde fluorescente (GFP), teniendo así la señal de salida. Dentro del módulo lógico se encuentran acomodados terminadores transcripcionales asimétricos que interrumpen el flujo de la RNA polimerasa sólo cuando se encuentran en una orientación particular. Mediante la acción de las integrasas es posible cambiar dicha orientación, y por lo tanto la actividad del terminador, puesto que los elementos lógicos se encuentran flanqueados por sitios de reconocimiento BP dispuestos en dirección contraria. En esta configuración son las integrasas quienes tienen el papel de las señales de control y cada compuerta es controlada por un par de ellas. Así, por ejemplo, la compuerta XOR sólo tendrá una señal alta cuando una u otra de las integrasas de control se expresen, de lo contrario, el terminador transcripcional se encontrará en su estado activo (Figura 1 panel B). Se utilizó la misma estrategia para generar las 6 funciones lógicas diferentes. Nótese que para generar la función lógica NAND fue necesaria la inclusión de un promotor constitutivo.



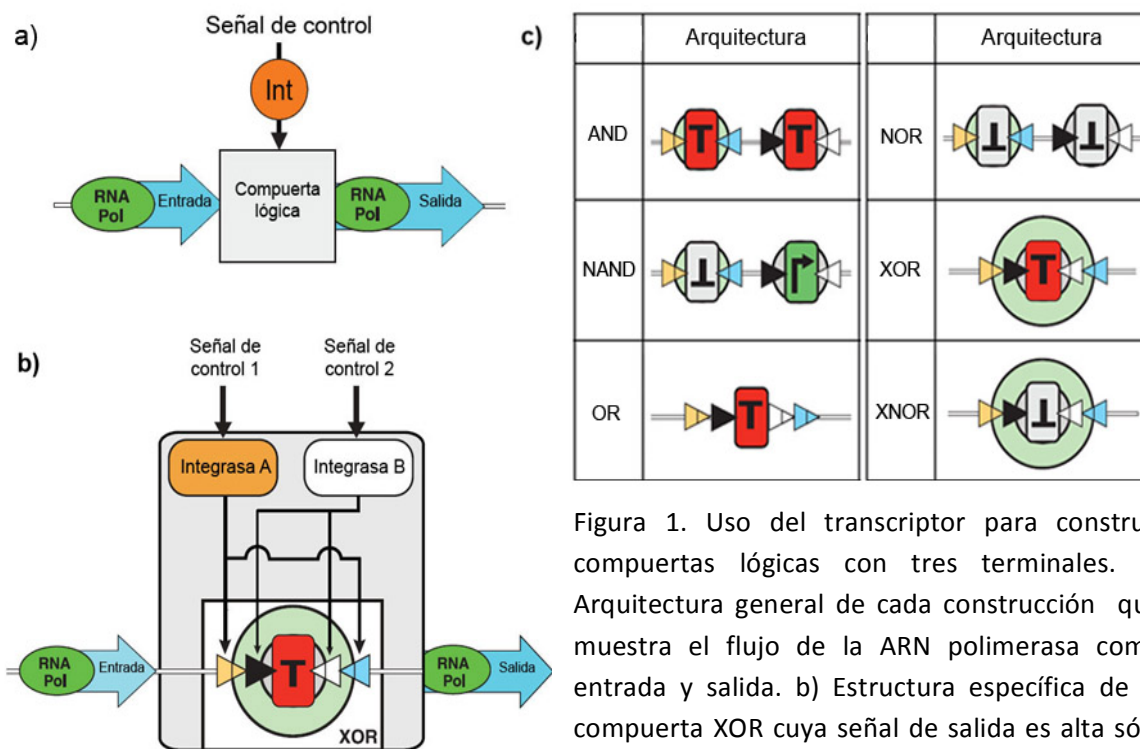


Figura 1. Uso del transcriptor para construir compuertas lógicas con tres terminales. a) Arquitectura general de cada construcción que muestra el flujo de la ARN polimerasa como entrada y salida. b) Estructura específica de la compuerta XOR cuya señal de salida es alta sólo cuando las señales de control son diferentes. c) Configuración de cada compuerta.

Si bien las compuertas difieren en su topología lógica, todas conservan el mismo número de moléculas involucradas. La señal de salida se puede observar mediante la expresión de la GFP, cuyo RNA se transcribe a una tasa dependiente del promotor constitutivo siempre y cuando el estado del módulo lógico lo permita. La expresión de las integrasas, por otro lado, es controlada mediante promotores inducibles. Los experimentos se realizaron utilizando las integrasas TP901-1 y Bxb1. La primera integrasa es una enzima de 485 aminoácidos aislada del fago TP901-1, la segunda integrasa consta de 500 aminoácidos y viene del micobacteriofago Bxb1 (Keravala *et al.*, 2006). Ambas se expresan bajo el control de los promotores inducibles por arabinosa (ara) y anhidrotetraciclina (aTc), respectivamente. De esta manera las señales de control de las compuertas se encuentran bajo el mando del experimentador, quien puede adicionar la molécula en cuestión.

## ❖ Modelo de los promotores inducibles

Se utilizaron dos promotores inducibles: el promotor  $P_{BAD}$  que se activa con presencia de arabinosa y el promotor  $P_{Tet}$  que es inducido mediante el análogo de tetraciclina. La unión de los ligandos a sus factores de transcripción y la unión de los factores de transcripción al ADN son modeladas para estos dos promotores ( $P_{BAD}$  y  $P_{Tet}$ ) (Tamsir, Tabor y Voigt, 2011). Para la unión del ligando a su factor de transcripción en el equilibrio tenemos:

$$C = C_0 \frac{L^n}{K_d^n + L^n} \quad 1$$

donde  $C$  denota la concentración del factor de transcripción unido,  $C_0$  es la concentración total del factor de transcripción,  $L$  es la concentración del ligando,  $K_d$  es la asociación constante, y  $n$  es la cooperatividad. Por la ley de conservación de masas, la concentración del factor de transcripción libre  $C_F$  es

$$C_F = C_0 - C \quad 2$$

La unión de los factores de transcripción a sus promotores se modela en términos de fracción de promotores activos. Para ello se consideran todos los posibles estados de los promotores y aquellos estados en los que existe transcripción. Los estados posibles se esbozan en la Figura 2. La probabilidad de que cada promotor se encuentre en su forma de complejo abierto o activo  $P$  se describe a continuación:

$$P_{BAD} = \frac{K_1 + K_2 C}{1 + K_1 + K_2 C + K_3 C_F} \quad 3$$

$$P_{Tet} = \frac{K_1}{1 + K_1 + 2K_2 C_F + K_2^2 C_F^2} \quad 4$$

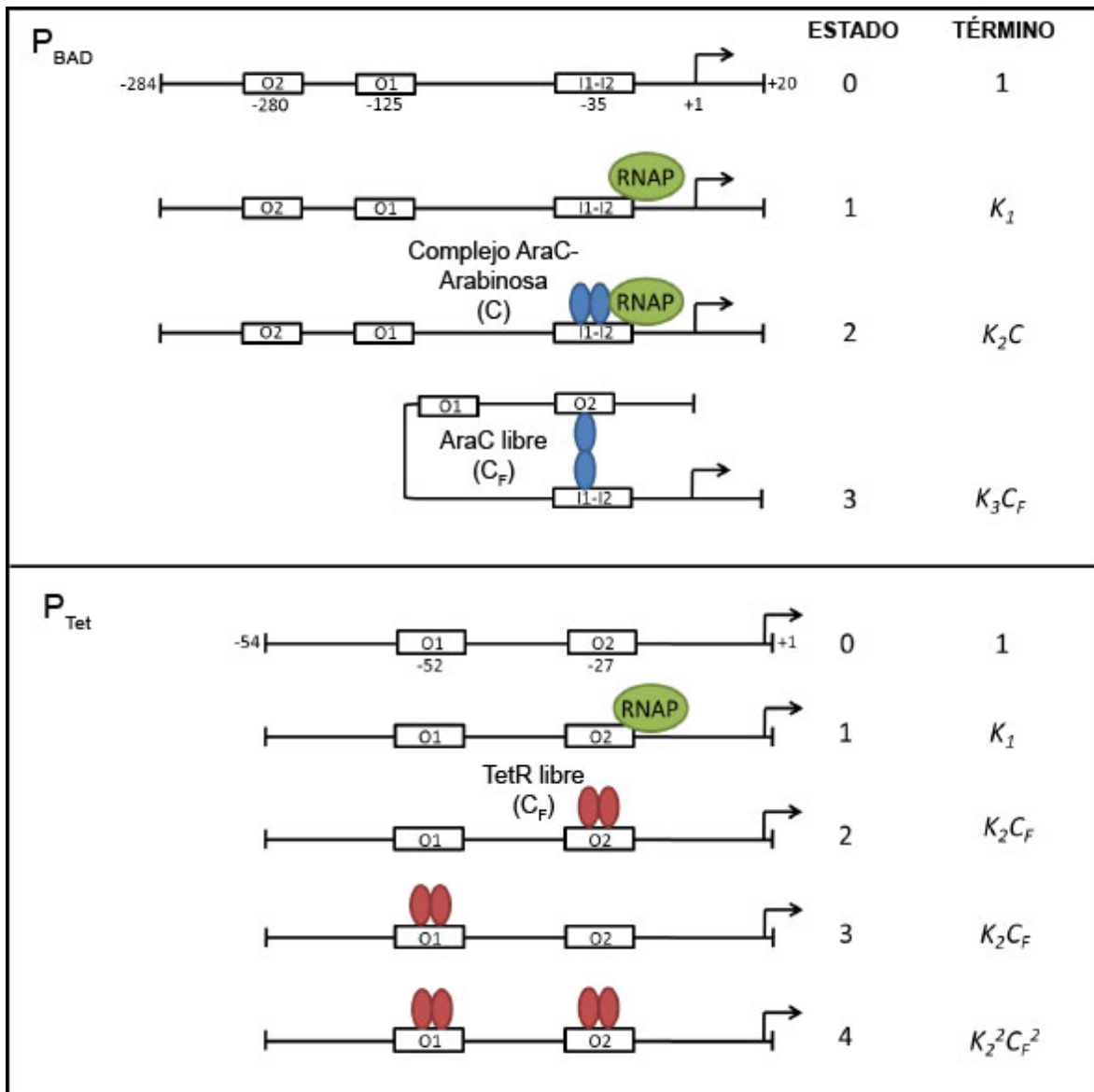


Figura 2. Estados de unión posibles para los promotores  $P_{BAD}$  y  $P_{Tet}$ . Los números describen la ubicación de los sitios del operador y los límites del promotor. Los términos con los que se representa cada estado en las funciones se indican hacia la derecha de cada diagrama.

Los parámetros de las funciones de probabilidad de apertura en los promotores se muestran en la Tabla 1 y fueron obtenidos previamente mediante el ajuste a funciones de transferencia experimentales y mediante modelos de los promotores en el equilibrio (Grigороva *et al.*, 2006).

Parámetro	$P_{BAD}$	$P_{Tet}$
$K_d$	0.09 mM	1.2 mM
$n$	2.8	1.0
$K_1$	0.009	350
$K_2$	37.5	160
$K_3$	3.4	

Tabla 1. Parámetros para los modelos de los promotores  $P_{BAD}$  Y  $P_{Tet}$ .

### ❖ Modelo determinista

El modelo determinista tiene como objetivo capturar la dinámica de una red de interacciones bioquímicas mediante una descripción matemática del fenómeno, de esta manera se puede analizar y simular su comportamiento para responder cuestiones específicas. Para ello se usó un conjunto de ecuaciones diferenciales ordinarias (ODEs) no lineales.

El sistema que concierne a este estudio se puede reducir a una red de interacción de seis especies químicas diferentes por cada compuerta: ARNm de la integrasa A, ARNm de la integrasa B, integrasa A, integrasa B, ARNm de la proteína verde fluorescente (GFP) y la proteína GFP. Para obtener el conjunto de ODEs, cada especie química se asignó a una sola variable de estado,  $X(t)$  = concentración de la especie X. La colección de valores de todas estas variables ( $X_1(t)$ ,  $X_2(t)$ ... $X_6(t)$ ) en un punto específico de tiempo constituyen el estado del sistema. Posteriormente, para cada especie molecular, se concibió una ecuación diferencial que describe el cambio en su concentración a través del tiempo como consecuencia de su interacción en las otras especies dentro de la red. Siguiendo este esquema se llegó al conjunto de ecuaciones diferenciales expuestas en la Tabla 2. Nótese que adicionalmente de las seis ecuaciones diferenciales, se incluyen un par de expresiones algebraicas. Estas expresiones describen la proporción de los sitios que son el blanco de las integrasas en el ADN y que han sido invertidos. Su forma es consecuencia de que, experimentalmente, al graficar el porcentaje de sitios invertidos contra condiciones relacionadas con la densidad de integrasas en el medio, se obtienen curvas sigmoideas que pueden ser aproximadas a funciones de tipo Hill (Bonnet *et al.*, 2013; Zhang *et al.*, 2008).

Especie molecular	Expresión
ARNm de integrasa A	$\frac{dIntAm}{dt} = k_{tra} \cdot P_{BAD} - k_{dam} \cdot IntAm$
Integrasa A	$\frac{dIntA}{dt} = k_{tna} \cdot IntAm - k_{dap} \cdot IntA$
ARNm de integrasa B	$\frac{dIntBm}{dt} = k_{trb} \cdot P_{Tet} - k_{dbm} \cdot IntBm$
Integrasa B	$\frac{dIntB}{dt} = k_{tnb} \cdot IntBm - k_{dbp} \cdot IntB$
Estado invertido del blanco de la integrasa A	$X_{LRa} = \frac{IntA^n}{IntA^n + a^n}$
Estado invertido del blanco de la integrasa B	$X_{LRb} = \frac{IntB^n}{IntB^n + b^n}$
ARNm de Gfp	$\frac{dGFPm}{dt} = k_{trg} \cdot pSC101 \cdot f(X_{LRa} \cdot X_{LRb}) - k_{dgm} \cdot GFPm$
Gfp	$\frac{dGFP}{dt} = k_{tng} \cdot GFPm - k_{dgp} \cdot GFP$

Tabla 2. Sistema de ODEs que describe la dinámica general de cada compuerta lógica. Las k's representan las tasas de síntesis o degradación, dependiendo del signo del término correspondiente.

La tasa a la que sucede cada reacción (de síntesis o de degradación) fue representada mediante una ley cinética particular, y cada una de ellas se asocia con una tasa constante o parámetro. En la Tabla 3 se incluye la totalidad de los parámetros utilizados en el sistema de ecuaciones diferenciales. Estos parámetros son, en su mayoría, aproximaciones desde sistemas de expresión similares puesto que en la literatura es muy limitada la presencia de las constantes relacionadas a la

acción de las integrasas. En otros casos fue posible estimar parámetros a partir de las vidas medias reportadas para las moléculas involucradas (valores marcados con asterisco en la Tabla 3).

Constante	Indicación	Valor (nM/s)	Referencias
$k_{tra}$	Transcripción del gen de la integrasa A	$2.48 \times 10^{-2}$	Kierzek <i>et al.</i> (2001),
$k_{dam}$	Degradación del ARNm de la integrasa A	$5.3 \times 10^{-3}$ *	Bernstein <i>et al.</i> (2002)
$k_{trb}$	Transcripción del gen de la integrasa B	$2.48 \times 10^{-2}$	Kierzek <i>et al.</i> (2001)
$k_{dbm}$	Degradación del ARNm de la integrasa B	$5.3 \times 10^{-3}$ *	Bernstein <i>et al.</i> (2002)
$k_{tna}$	Traducción del ARNm de la integrasa A	$2.6 \times 10^{-2}$	Watson <i>et al.</i> (2003)
$k_{dap}$	Degradación de la integrasa A	$4.5 \times 10^{-4}$	Pratt <i>et al.</i> (2002)
$k_{tnb}$	Traducción del ARNm de la integrasa B	$2.6 \times 10^{-2}$	Watson <i>et al.</i> (2003)
$k_{dbp}$	Degradación de la integrasa B	$4.5 \times 10^{-4}$	Pratt <i>et al.</i> (2002)
$k_{trg}$	Transcripción del gen de la GFP	$2.48 \times 10^{-2}$	Kierzek <i>et al.</i> (2001)
$k_{dgm}$	Degradación del ARNm de la GFP	$9.95 \times 10^{-3}$ *	Yuichi Taniguchi <i>et al.</i> (2010)
$k_{tng}$	Traducción del ARNm de la GFP	$2.6 \times 10^{-2}$	Watson <i>et al.</i> (2003)
$k_{dgp}$	Degradación de la GFP	$8.2 \times 10^{-4}$ *	Longo y Jeff Hasty (2006)
<b>n</b>		2	Bonnet <i>et al.</i> (2013)
<b>a</b>	Ajuste a la función de Hill para la integrasa A	$3 \times 10^{-3}$	Bonnet <i>et al.</i> (2013)
<b>b</b>	Ajuste a la función de Hill para la integrasa B	1.5	Bonnet <i>et al.</i> (2013)
<b>pSC101</b>	Número de plásmidos del vector de expresión de la GFP	7	Peterson y Philips (2008)

Tabla 3. Parámetros cinéticos usados en el modelo determinista.

En la Tabla 2 podemos observar que dentro del término de producción del ARNm de la GFP aparece una función genérica indefinida. Esta función representa la operación lógica y es específica para cada compuerta lógica, por lo que se generaron seis expresiones lógicas. Las expresiones relacionan cada operación lógica con las señales de control constituidas por las integrasas y permiten calcular la fracción de sitios transcripcionalmente activos que generan una señal de salida alta. Estas funciones se obtuvieron utilizando las leyes de la cinética enzimática. La Tabla 4 presenta la colección de funciones lógicas.

Operación	Expresión
AND	$f_{(ON)} = \frac{XlrA \cdot XlrB}{1 + XlrA + XlrB + XlrA \cdot XlrB}$
NAND	$f_{(ON)} = \frac{1 + XlrA + XlrB}{1 + XlrA + XlrB + XlrA \cdot XlrB}$
OR	$f_{(ON)} = \frac{XlrA + XlrB + XlrA \cdot XlrB}{1 + XlrA + XlrB + XlrA \cdot XlrB}$
NOR	$f_{(ON)} = \frac{1}{1 + XlrA + XlrB + XlrA \cdot XlrB}$
XOR	$f_{(ON)} = \frac{XlrA + XlrB}{1 + XlrA + XlrB + XlrA \cdot XlrB}$
XNOR	$f_{(ON)} = \frac{1 + XlrA \cdot XlrB}{1 + XlrA + XlrB + XlrA \cdot XlrB}$

Tabla 4. Funciones lógicas para el control de la señal de salida.

## ❖ Modelo estocástico

El sistema de ecuaciones diferenciales representa una colección de reacciones ocurriendo de manera simultánea dentro de un volumen determinado. Sin embargo, las suposiciones de simplificación que permiten utilizar esta aproximación dejan de ser válidas cuando el número de moléculas es suficientemente pequeño, y entonces las reacciones involucradas ocurren en un orden azaroso y asincrónico. En estos casos es conveniente describir el sistema en cuestión a través de la probabilidad de estado de cada variable considerando el número de moléculas en lugar de usar concentraciones diferenciables en el tiempo. La descripción probabilística considera el ruido intrínseco de los sistemas de expresión genética y se refleja por la naturaleza estocástica de la reacción elegida en cada instante y por el tiempo que transcurre entre cada reacción (Gillespie, 1992). El algoritmo de Gillespie es una manera de implementar consistentemente la descripción probabilística de los sistemas biológicos y fue utilizada en este trabajo para construir el modelo estocástico.

El algoritmo, entonces, consiste en la iteración de los siguientes pasos:

1. Iniciar con el tiempo  $t=t_0$  y el estado del sistema  $x=x_0$ .
2. Con el sistema en el estado  $x$  y el tiempo  $t$ , evaluar todos los términos de las variables  $a_j(x)$  y su suma  $a_0(x)$ .
3. Generar valores para  $\tau$  y  $j$  de acuerdo a las dos expresiones 7 y 8 (Ver apéndice A).
4. Efectuar la siguiente reacción elegida reemplazando  $t \rightarrow t + \tau$  y  $x \rightarrow x + v_j$ .
5. Guardar las actualizaciones del sistema  $(x,t)$ . Regresar al paso 2 y continuar hasta que se exceda el tiempo de simulación.

Con el objetivo de implementar el algoritmo de Gillespie se definió cada compuerta lógica como un conjunto de posibles reacciones con cierta probabilidad de ocurrencia. Para un instante de tiempo particular, todas las reacciones tienen una probabilidad individual de acontecer; a dicha probabilidad se le denomina propensión.



Así, cada propensión  $a_j$  es proporcional a la probabilidad de que su correspondiente reacción  $R_j$  ocurra, y ello está dado por la función de densidad de  $j$ ,  $a_j(x)/a_0(x)$  (Ver Apéndice A). A este término se refiere comúnmente como propensión. Una propensión es el producto de dos partes: la tasa de reacción y el número de reacciones posibles dado el número de moléculas disponibles, y por lo tanto, un sistema posee tantas propensiones como reacciones distintas. Las propensiones pueden ser definidas a partir del modelo determinista si se considera el volumen de reacción para, entonces, expresar las tasas de reacción como número de moléculas en lugar de unidades de concentración. Siguiendo este orden de pensamiento, se estableció el conjunto de reacciones derivadas del procesamiento de información lógica en cada compuerta y se muestran en la Tabla 5, con sus respectivas propensiones de ocurrencia. De manera global son 14 reacciones y la propensión total es la suma de las individuales.

Número	Reacción	Propensión
$a_1$	$ADN \rightarrow IntAm$	$PBAD k_{tra}$
$a_2$	$IntAm \rightarrow \emptyset$	$IntAm k_{dam}$
$a_3$	$IntAm \rightarrow IntA$	$IntAm k_{tna}$
$a_4$	$IntA \rightarrow \emptyset$	$IntA k_{dap}$
$a_5$	$ADN \rightarrow IntBm$	$PTet k_{trb}$
$a_6$	$IntBm \rightarrow \emptyset$	$IntBm k_{dbm}$
$a_7$	$IntBm \rightarrow IntB$	$IntBm k_{tnb}$
$a_8$	$IntB \rightarrow \emptyset$	$IntB k_{dbp}$
$a_9$	$X_{BPa} \rightarrow X_{LRa}$	$IntA^n / IntA^{n+a^n}$
$a_{10}$	$X_{BPb} \rightarrow X_{LRb}$	$IntB^n / IntB^{n+b^n}$
$a_{11}$	$ADN \rightarrow GFPm$	$f(X_{LRa}, X_{LRb}) k_{trg}$
$a_{12}$	$GFPm \rightarrow \emptyset$	$GFPm k_{dgm}$
$a_{13}$	$GFPm \rightarrow GFP$	$GFPm k_{tng}$
$a_{14}$	$GFP \rightarrow \emptyset$	$GFP k_{dgp}$

Tabla 5. Propensiones del algoritmo estocástico. Ver la Tabla 2 para la notación de los términos.

### ❖ Circuitos extendidos

Los modelos determinista y estocástico descritos anteriormente fueron establecidos para representar cada compuerta lógica. Sin embargo, es parte de los objetivos de este trabajo abrir la posibilidad de la concepción de circuitos de mayor complejidad. Por lo tanto, mediante la conexión de múltiples compuertas a las señales de entrada y entre sí, se construyeron tres circuitos: un semisumador, un sumador complejo y un decodificador 2x4.

La conexión entre compuertas lógicas implica que la señal de salida de un dispositivo funciona como la señal de entrada del siguiente dispositivo de la red. En este caso se tiene un sistema de señales de control mediadas por integrasas, por lo que la señal de salida de una compuerta permite la expresión de una integrasa única y funcionalmente ortóloga dentro de la red. La integrasa expresada funge entonces como señal de control para una compuerta subsiguiente.

Todas las reacciones para un mismo circuito ocurren dentro de un volumen común, de ahí la importancia de que cada señal se represente por una integrasa diferente. Los modelos para las compuertas aisladas funcionan en base a los parámetros determinados de manera experimental para dos integrasas: Bxb1 y TP901-1. Sin embargo, no se tiene acceso a los parámetros necesarios para representar la dinámica y la acción enzimática de integrasas adicionales. En consecuencia, para la construcción de los tres circuitos mencionados anteriormente, se asumió que la dinámica y acción de otras enzimas hipotéticas son similares y se utilizó el mismo conjunto de parámetros.

### ❖ Análisis de simulación

Tanto para el modelo estocástico como para el determinista se tiene una red de moléculas que interaccionan sin involucrar lazos de retroalimentación. Es decir, cada molécula específica altera la dinámica en nodos posteriores de la red y no nodos previos. Como resultado, la concentración o número de moléculas para cada variable llega a un punto estable en algún momento de la simulación. Por lo tanto, en todos los casos el análisis de los resultados fue realizado sólo cuando se aseguró la condición del estado estable.

Al igual que para los sistemas electrónicos, el objetivo del procesamiento de los datos fue la digitalización de la información. Ello se logró mediante la distinción inequívoca de los estados encendido y apagado en las señales de salida de cada dispositivo. En el modelo determinista se midió la señal de salida en términos de concentración de moléculas. Por otro lado, el modelo estocástico arroja número de moléculas como señal de salida; adicionalmente, cada dispositivo se caracterizó mediante la media ( $\mu$ ) y la desviación estándar ( $\sigma$ ). También fue de interés determinar la magnitud de la variación de las señales de salida, y para ello se calculó el coeficiente de variación dado por  $Cv = \sigma/\mu$ .

# RESULTADOS Y DISCUSIÓN

---

## ❖ Promotores inducibles

La primer parte del modelo para cada compuerta consiste en la representación de los promotores inducibles. El promotor  $P_{BAD}$  se activa con presencia de arabinosa y el promotor  $P_{Tet}$  es inducido mediante el análogo de tetraciclina. Estos promotores regulan la expresión de las integrasas A (Bxb1) y B (TP901-1), respectivamente.

Los promotores utilizados contienen secuencias operadoras, cada una de las cuales es reconocida por más de un tipo de factores de transcripción y represores con distintas afinidades (Zhang y Tianshaou, 2014). La unión de estos factores a sus operadores da lugar a múltiples estructuras, algunas de las cuales son transcripcionalmente activas mientras que otras no lo son. En este caso, la presencia de los inductores permite la formación de complejos con las moléculas represores dejando así a las secuencias operadoras en su estado abierto, lo que posibilita la formación del complejo transcripcional y el inicio del proceso de expresión (Zhang y Tianshaou, 2014) (ver Figura 2). Para representar este fenómeno se utilizó un modelo enfocado en determinar la probabilidad con la que los operadores se encuentran en su estado abierto en función de la concentración de su molécula inductora. Aunque existen algunas estrategias para representar la transición de los promotores de su estado activo a inactivo y viceversa, los modelos teóricos usados para este trabajo omiten este efecto puesto que son útiles para calcular los niveles de expresión en el estado estable según las condiciones específicas de concentración de las moléculas regulatorias (Bintu et al., 2005). La Figura 3 muestra las dinámicas de los promotores estudiados. Los cálculos se realizaron considerando que el valor de concentración de los represores se encuentra en órdenes de magnitud cercanos a  $1 \times 10^{-6}$  M. Las gráficas de la derecha corresponden al promotor  $P_{BAD}$ , las de la izquierda, al promotor  $P_{Tet}$ . Se advierte que la probabilidad de apertura aumenta con la concentración del

inductor, puesto que el inductor favorece a las estructuras transcripcionalmente activas de los promotores. Los ejes y de las gráficas de la primera línea presentan una escala de 0 a 1 porque denotan la probabilidad de apertura del promotor correspondiente.

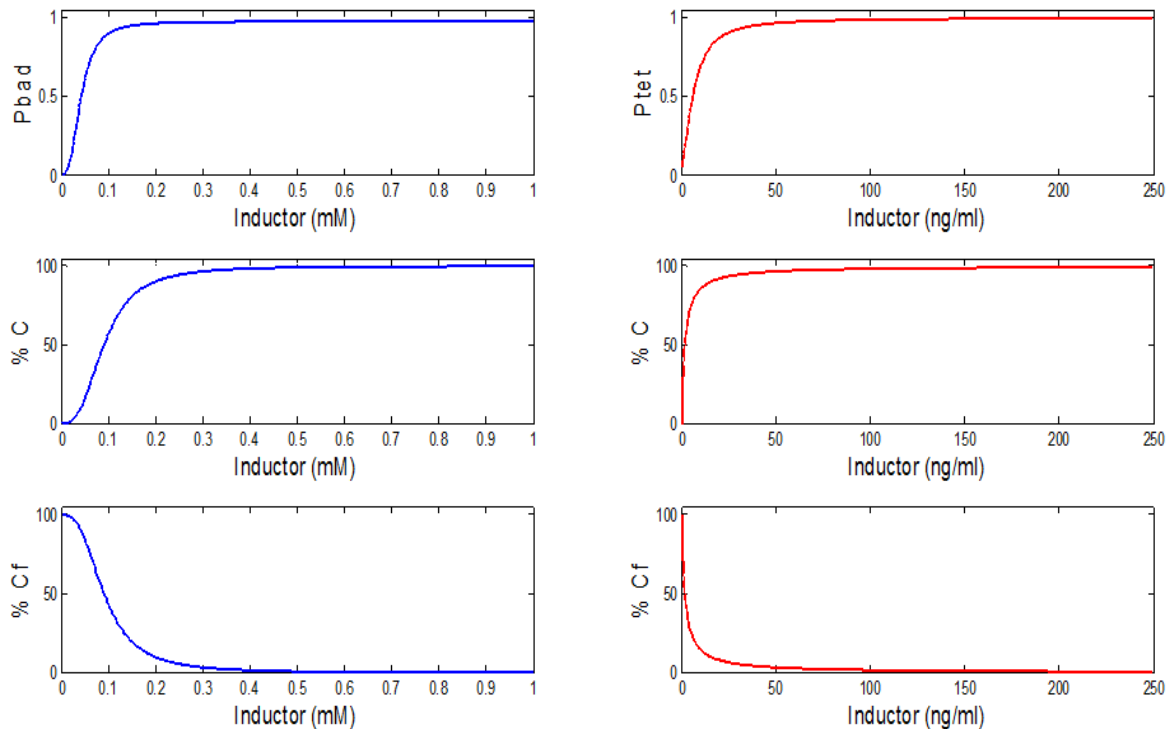


Figura 3. Dinámica de los promotores  $P_{BAD}$  y  $P_{Tet}$ . Se grafica la relación entre la concentración del inductor y la probabilidad de apertura del promotor (primera fila), el porcentaje de represores en complejo con su operador (segunda fila) y el porcentaje de represores libres (tercera fila). En azul se tiene la dinámica de  $P_{BAD}$  y el rojo es para  $P_{Tet}$ .

Experimentalmente, el promotor  $P_{BAD}$  se induce bajo condiciones normales utilizando concentraciones de 1% p/v de arabinosa (Tamsir, Tabor y Voigt 2011); esta concentración equivale a 60 mM. El modelo indica que el inductor está totalmente activo en concentraciones superiores a 1mM, por lo que la predicción coincide al menos con el límite práctico de concentración. Por otro lado, para el caso de  $P_{Tet}$ , durante experimentos de expresión frecuentemente se reporta su inducción utilizando concentraciones de alrededor de 200 ng/ml de anhidrotetraciclina (Carroll *et al.*, 2005) y, de igual manera, el modelo indica que a esta concentración la probabilidad de apertura es prácticamente del 100%. La forma sigmoidea de las curvas dosis-respuesta se ha observado antes tanto para el promotor inducible por arabinosa (Guzman *et al.*, 2005) como para el inducible por

tetraciclina (Loew et al., 2010) y las concentraciones de inductor a las que se presenta la saturación son consistentes con lo reportado.

#### ❖ Análisis determinista

Las señales de control están constituidas por las integrasas y su expresión representa la primera parte del modelo y como se mencionó anteriormente, la concentración de las integrasas está sujeta a la presencia del inductor que permite la activación del promotor que controla su expresión; este proceso tiene como resultado la síntesis del ARNm correspondiente de cuya traducción deriva la proteína. La dinámica de la sucesión de eventos de síntesis se estudió mediante las primeras cuatro ecuaciones diferenciales de la Tabla 2. Los resultados del análisis de la integrasa A se pueden apreciar en la Figura 4. Las gráficas en rojo, azul y verde representan los niveles de activación del promotor y la expresión del ARNm y de la proteína, respectivamente. Se obtuvieron predicciones similares para la integrasa B. Este tipo de modelo permite estudiar los sistemas biológicos de manera precisa en el estado estable de las variables, de tal manera que se puede estimar la ocurrencia de interacciones celulares bajo estas condiciones (Zhang y Tianshou, 2014). Consecuentemente podemos observar que cada una de las variables involucradas en la aparición de las integrasas alcanzan su estado estable, cuyo valor depende de los parámetros del sistema. Existen periodos transitorios antes de que el máximo nivel de expresión sea alcanzado. El ARNm tiene una dinámica bastante rápida y el estado estable se alcanza a un valor cercano a los 15 minutos después de que el promotor se encuentre en su forma totalmente activa; el mismo proceso para las proteínas toma un poco menos de 200 minutos. En este punto la literatura es bastante variada, siendo que algunas simulaciones deterministas reportan transitorios para la dinámica de moléculas de ARNm del orden de los 400 minutos (Hoffman *et al.*, 2008) o periodos de 20 o 30 minutos para moléculas proteicas (Arpino *et al.*, 2013), siendo que el periodo de estabilización está relacionado al juego de parámetros que se utilicen para las reacciones de síntesis y degradación. El transitorio de 200 minutos aquí obtenido indica que se requieren unas 7 generaciones para alcanzar el punto de equilibrio, lo que es concebible dado que durante el proceso de división celular, las biomoléculas se reparten entre las células hijas y ello ha sido reportado con datos experimentales (Huh y Paulsson, 2010).

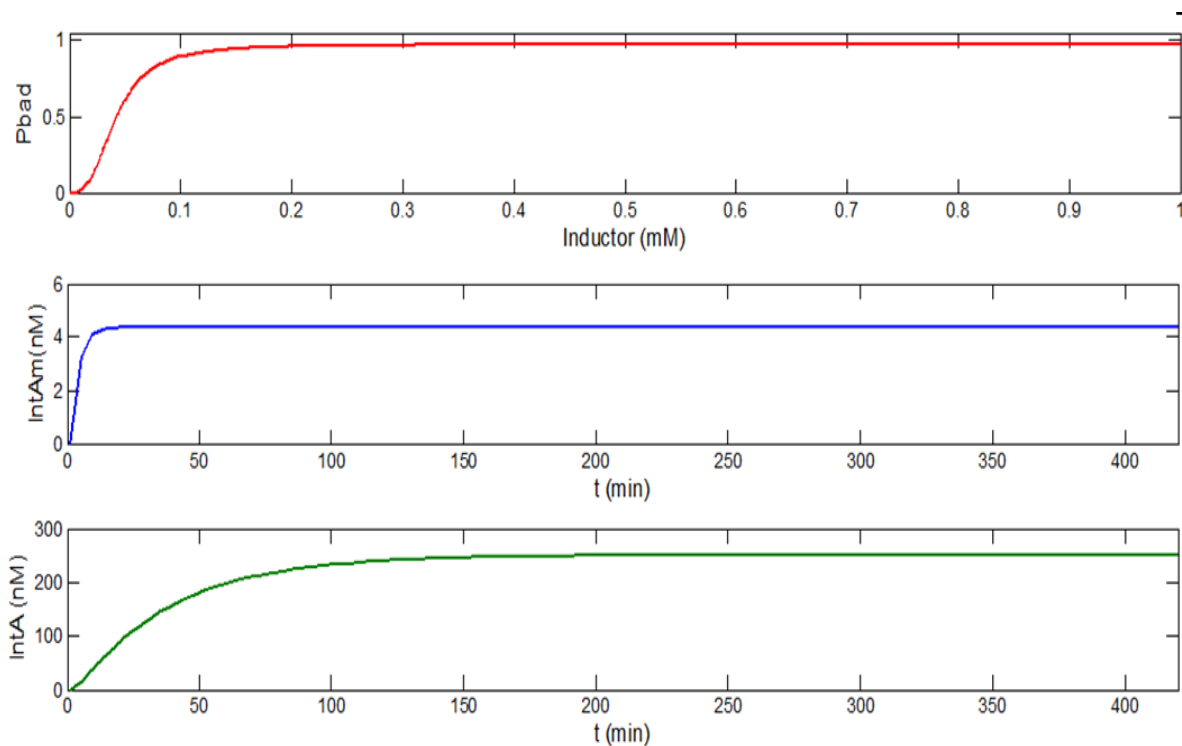


Figura 4. Dinámica de la expresión de la integrasa A. Se muestra la evolución de la probabilidad de apertura del promotor  $P_{BAD}$  como función de la concentración del inductor (rojo), y el cambio en los niveles del ARNm de la integrasa (azul) y la proteína (verde) a través del tiempo.

El punto de equilibrio del sistema es de suma importancia porque indica la concentración de metabolitos dentro de la red, y éstos deben de encontrarse en los rangos biológicos para una representación de la realidad razonablemente fiel. Estudios de transcriptoma y proteoma en *E. coli* concluyen que la abundancia promedio del ARNm de 137 genes se encuentra en el rango de 0.05 a 5 copias por célula, y todos guardan una correlación significativa con el nivel de expresión de las proteínas correspondientes considerando el número de genes (Tanigucho *et al.*, 2010); algunos mensajeros llegan a las 10 moléculas por células. El modelo de expresión de las integrasas predice una concentración de 4.2 nM del ARNm de la integrasa A (Bxb1) en su punto de equilibrio. El ARNm mensajero de esta enzima es de 1560 nucleótidos (Kim *et al.*, 2003) y, si consideramos un volumen celular de  $1\mu\text{m}^3$ , tenemos que se generan 2.5 copias de mensajero en el equilibrio y ello lo clasificaría dentro del tipo de genes de baja expresión (Tanigucho *et al.*, 2010). Lo mismo ocurre para la integrasa B.

El mismo análisis se realizó para la concentración de las integrasas. Existe poca documentación sobre la cuantificación de la expresión *in vitro* o *in vivo* de las integrasas, ya que la mayor parte de su caracterización es de naturaleza funcional, y la información cuantitativa no representa las condiciones celulares normales. Por ejemplo, Stoll y colaboradores (2002) reseñan el análisis de extractos de la integrasa TP901-1 recombinante obtenidos de cepas de *E. coli* y reporta concentraciones de 7.4 mg/ml (134  $\mu$ M), lo que equivale a un aproximado de 81,030 moléculas de integrasa por célula. Sin embargo estos experimentos se realizaron mediante vectores de expresión de alto rendimiento (vector de expresión pET 11) destinados a obtener altos rendimientos proteicos, y por lo tanto no reflejan las condiciones que aquí se pretenden representar con el modelo *in silico*. Por otro lado, es posible auxiliarse nuevamente de estudios de proteoma en bacterias, específicamente en el microorganismo modelo *E. coli*. Dichos estudios indican que el número de copias de proteínas abarca un amplio rango de valores que va de 50 a  $10^5$  moléculas por célula, siendo que las proteínas de baja abundancia se presentan con un promedio cercano a las 500 moléculas (Ishihama *et al.*, 2008). El modelo determinista predice una expresión máxima de 255 nM en el punto de equilibrio, lo que equivale a 153.3 copias de integrasa A por célula. Según (Taniguchi *et al.*, 2010), esta proteína entraría en el rango de las moléculas de baja expresión, lo que era de esperar puesto que no representa ninguna función esencial para la célula.

Las integrasas responden bien a la activación de su expresión por los inductores; sin embargo representan apenas las señales de control del sistema. En el caso de la arquitectura del transcriptor, las señales de control cumplen el papel de las señales de entrada de cuyo estado depende la salida, y es ésta última quien determina si el dispositivo responde con el estado de encendido o apagado. Para evaluar las señales de salida se interrogó cada una de las compuertas de acuerdo a las diferentes combinaciones de estado de las señales de entrada utilizando las ecuaciones diferenciales restantes que describen la expresión de la Gfp a través de las distintas operaciones lógicas (Tabla 2). La evaluación puede ser ilustrada con el análisis de la compuerta AND, que constituye una de las compuertas más simples dentro de la electrónica digital. Ésta entregará una salida alta (1), dependiendo de los valores de las entradas, siendo este caso, al recibir solo valores altos. Si alguna o ambas entradas no son altas, entonces se mostrará un valor de salida bajo (0) (Mano, 1994). La lógica de la compuerta se resume en la tabla de verdad que aparece en el panel superior de la Figura 5. En la parte inferior, dentro de la misma figura, se expone la



interrogación de la compuerta AND según el modelo determinista mediante una serie de gráficas que responden al nivel de expresión en el tiempo de las moléculas involucradas. Siguiendo el esquema de la tabla de verdad, la dos primeras columnas de gráficas pertenecen a las señales de control del transcriptor (análogas a las señales de entrada desde la perspectiva de la lógica electrónica), integrasa A e integrasa B, respectivamente; la última columna es para la señal de salida, la proteína reportera Gfp. Desde una vista horizontal, cada fila de gráficas representa una de las cuatro combinaciones posibles para las señales de entrada (00, 10, 01 y 11) y la correspondiente respuesta en la señal de salida. Tratándose de dinámicas moleculares en el tiempo, el eje x sigue la evolución temporal en minutos, y el eje y indica la concentración en unidades de nM. Las dinámicas simuladas demuestran que el modelo es capaz de reproducir el procesamiento lógico de la información en términos de concentración de moléculas pues sólo cuando la expresión de las dos integrasas es alta se genera una expresión alta de Gfp. En el estado apagado o bajo (0), la concentración de Gfp es prácticamente nula mientras que en el estado encendido o alto (1) la proteína alcanza una concentración de ~142 nM y por lo tanto es fácil establecer un valor umbral, a nivel de compuerta individual, que permita discernir entre los estados prendido y apagado. Lo anterior hace referencia al rango dinámico del dispositivo, entendiéndose como la relación del máximo nivel especificado de un parámetro respecto valor perceptible mínimo (Mano, 2014), por lo que entre más alto sea el rango dinámico, mayor será la facilidad de juzgar una señal como alta o baja. En muchas de las construcciones biológicas de compuertas lógicas propuestas el rango dinámico es bastante bajo y ello representa una desventaja, especialmente cuando se desea ensamblar redes de múltiples puertas (Sayut, Niu y Sun, 2009).

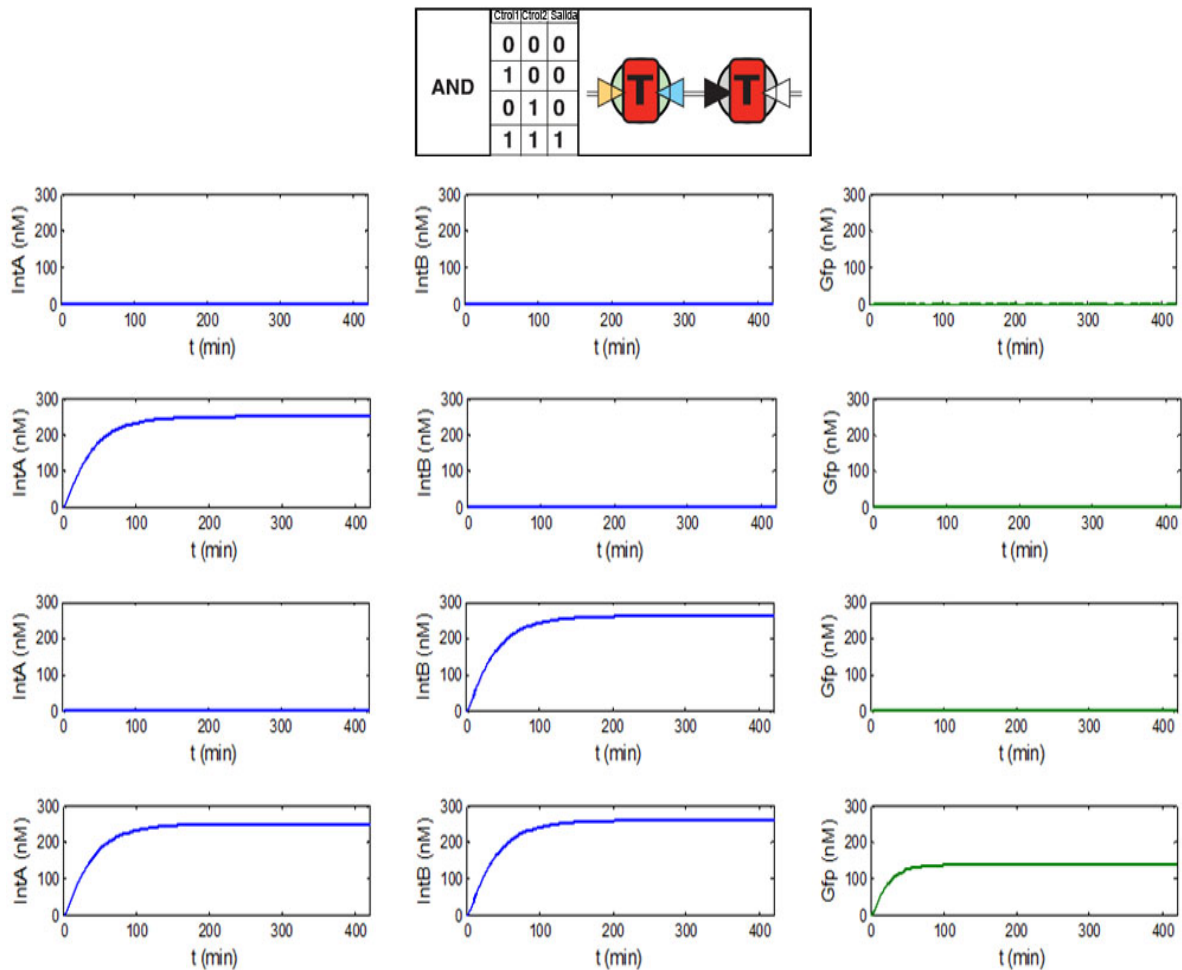


Figura 5. Reproducción de la tabla de verdad de la compuerta AND a través del modelo estocástico. El panel superior esquematiza la tabla de verdad para la compuerta AND y su arquitectura de construcción. Ordenadas de manera vertical, las primeras dos columnas de gráficas (en azul) indican la expresión de las integrasas A y B, respectivamente, como función del tiempo; la última columna (en verde) representa la expresión de la GFP. De manera horizontal, cada fila de columnas indica los cuatro estados posibles de las señales de entrada y su respuesta en la señal de salida.

Se realizó el análisis anterior para cada una de las compuertas y en la columna con las barras verde claro de la Figura 8 pueden apreciarse los resultados de los niveles de expresión normalizados para la GFP. En todos los casos en que las señales de salida, de acuerdo a cada operación lógica particular, deben ser bajas o comportarse como apagadas (0), la expresión predicha de GFP es prácticamente nula. También ocurre el comportamiento esperado cuando las señales de salida

deben ser altas (1); sin embargo, el nivel de expresión en el estado encendido varía entre algunas compuertas. El contraste más notorio ocurre con la compuerta AND, donde la expresión en el estado encendido es apenas el 25% de la expresión máxima observada en otras compuertas. Este fenómeno es efecto de las expresiones lógicas que determinan el procesamiento de información de cada compuerta. Estas expresiones se muestran en la Tabla 4 y fueron inspiradas en el análisis de cinética enzimática; en el numerador se posicionan los términos relacionados a los estados encendidos de una compuerta en específico, mientras que en el denominador están todos los términos representando los cuatro posibles estados (ambas integrasas, ninguna integrasa, integrasa A o integrasa B presentes). Siendo así, las expresiones determinan la fracción de sitios activos que permiten el flujo de la RNA polimerasa hacia el gen de la GFP, pero por su estructura, tanto el valor del numerador como el del denominador varían con cada estado de las señales de entrada (Sauro, 2011). Por ejemplo, la diferencia se hace evidente entre las compuertas NOR y OR. El numerador de la función lógica NOR tiene solamente un término y es igual a 1, por lo que el resultado máximo posible es de 1; por otro lado, el numerador de la función lógica OR es más complejo y el resultado máximo posible sólo se acerca a 1, dependiendo del valor que tomen el resto de los términos. Por la naturaleza de estas funciones se requerirían adicionar términos de ajuste y lograr así la homogenización de los valores máximos de expresión. Debido a ello, a diferencia del análisis de cada compuerta individual, es difícil establecer un nivel umbral de expresión para distinguir entre los estados apagado y prendido en el conjunto de compuertas vistas como un sistema integral.

#### ❖ **Análisis estocástico**

En muchos casos resulta conveniente la representación de las propiedades dinámicas de una red a través de ecuaciones diferenciales ordinarias; sin embargo, su uso se basa en algunas conjeturas. Una de ellas dice que las concentraciones químicas son funciones continuas en el tiempo, argumento válido si el número de moléculas de cada especie es suficientemente alto. La literatura indica que si el número total de moléculas de cualquier sustancia en particular es menor a un aproximado de 1000, entonces un sistema de ecuaciones diferenciales estocásticas es más apropiado (Rao *et al.*, 2002). Ya que en el sistema de las compuertas lógicas las proteínas se

presentan en el orden de centenas de moléculas y los ARNm en el orden de decenas o menor, es más apropiada la aproximación estocástica.

El modelo estocástico funciona con las probabilidades de que suceda una reacción determinada  $j$  dentro del volumen celular  $V$  en cada intervalo de tiempo infinitesimal. Esta probabilidad llamada propensión  $a_j$  depende de una constante, de manera que la probabilidad de tener la reacción  $j$  dado el número de moléculas  $x$  en un tiempo específico depende de una constante  $c_j$ , y así tenemos que  $a_j(x) = c_j x_i^n$ . Resulta que  $c_j$  se obtiene a partir de la constante  $k_j$  de la cinética química convencional determinista y es igual a  $k_j/V$  si los reactantes son moléculas diferentes (Gillespie, 1992). Por lo tanto, dado que los parámetros del modelo determinista generan resultados en concentraciones dentro del rango biológico experimental observado, podemos trasladarlos al algoritmo estocástico.

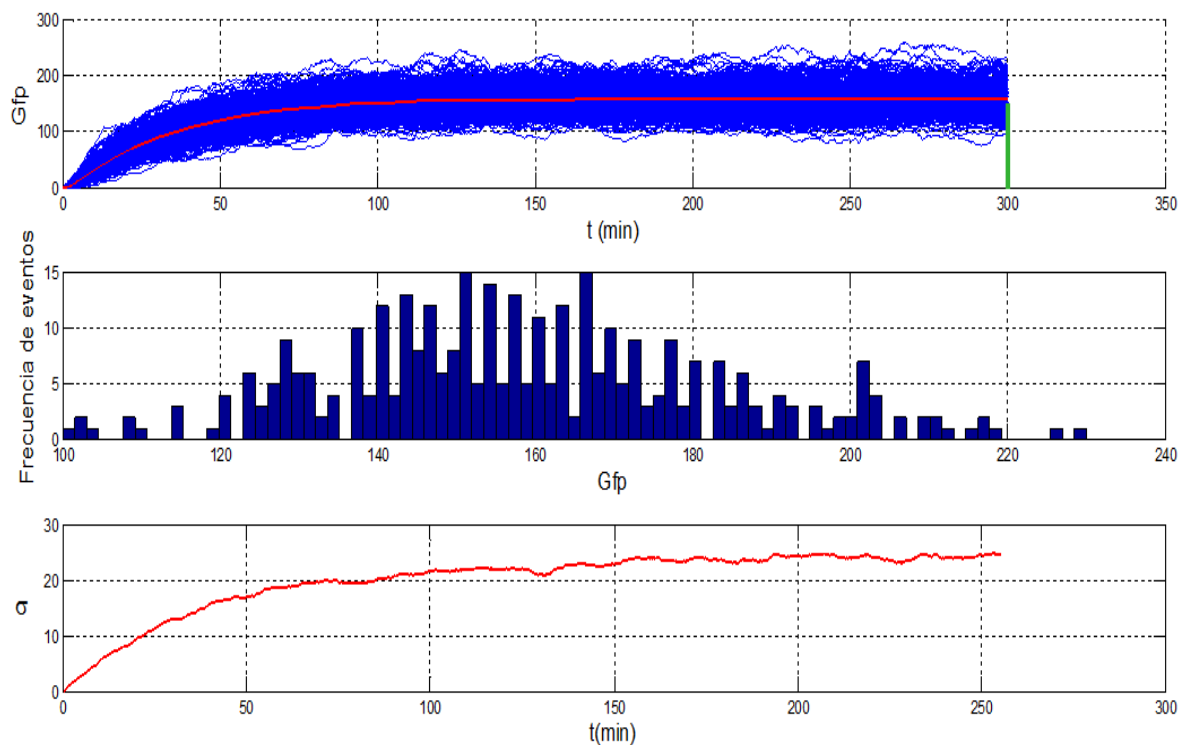


Figura 6. Simulación estocástica de la compuerta NAND. En el panel superior se trazan las trayectorias del nivel de expresión de GFP de 500 células (líneas azules), la predicción del modelo estocástico para la misma compuerta (línea roja) y el instante de tiempo en que se registraron los valores de las trayectorias (línea verde vertical). En el panel intermedio se encuentra el histograma construido a partir del registro de las trayectorias en el tiempo indicado. En el panel inferior se grafica la desviación estándar del número de moléculas de GFP para las 500 simulaciones a través del tiempo.

Por lo anterior, lo siguiente fue explorar el sistema de compuertas desde el punto de vista probabilístico en el ambiente celular. Para ello se evaluaron las funciones lógicas de las compuertas utilizando el algoritmo estocástico. El tipo de resultados que se obtienen de dicha exploración se muestra en la Figura 6. En la primera gráfica se observa la evolución del nivel de expresión de la GFP en el tiempo como resultado del conjunto de reacciones y propensiones del sistema de la compuerta NAND. Las señales de entrada para este caso fueron bajas o en estado apagado (00). Las trayectorias azules denotan el nivel de GFP en cada una de las células simuladas; la línea roja representa la predicción del modelo determinista; la línea verde vertical indica el punto en el tiempo en que se registró el número de moléculas de cada trayectoria y con esos datos se construyó el histograma del panel intermedio de la figura. Por último, en el panel inferior se grafica la desviación estándar del conjunto de simulaciones a través del tiempo, y demuestra que la variación máxima es 25.2 moléculas con respecto a la media. Lo mismo se realizó para cada una de las compuertas a través de todas las combinaciones de los estados de entrada posibles, y en la Figura 8 se resumen los resultados mediante las gráficas de barra color verde oscuro de la última columna, mismas que se interpretan como el nivel de expresión promedio de 500 simulaciones con sus respectivas barras de error.

Se esperaba que los niveles de expresión arrojados por el modelo determinista coincidieran con el promedio de las realizaciones generadas a partir del modelo estocástico (Stamatakis y Mantzaris, 2009). Esto ocurre para todos los casos a excepción de los estados encendidos de las compuertas AND y OR. Como se argumentó en el apartado del análisis determinista, estas compuertas generan salidas bajas en el estado encendido debido a las funciones lógicas que procesan la información de las señales de entrada. Sin embargo, utilizando el modelo estocástico este problema desaparece pues aparentemente el escenario probabilístico homogeniza la respuesta de las funciones lógicas y permite una señal de salida uniforme. Por lo tanto, el resultado es más consistente entre dispositivos cuando se considera el efecto aleatorio del número pequeño de moléculas. Otros estudios indican también que se obtienen resultados más realistas o bien que se identifican propiedades emergentes en casos similares donde el número de moléculas es crucial (Twycross *et al.*, 2010; Shnerb *et al.*, 2000).

Adicionalmente, el rango dinámico de las señales de salida en las compuertas AND y OR (y por lo tanto en el conjunto de compuertas) es mucho más conveniente en el contexto de reacciones como eventos aleatorios. El incremento y homogenización del rango dinámico es de gran importancia porque permite disminuir la probabilidad de obtener señales de salida apagado/prendido equivocadas o discordantes con las señales de entrada (Guet et al., 2002). Algunos investigadores han conseguido aumentar los rangos dinámicos en otras compuertas biológicas utilizando la función NOT de manera previa a la señal de salida (Tamsir, Tabot y Voigt, 2011), o alterando la secuencia y estructura de las partes biológicas mediante evolución sitio-dirigida (Moon et al., 2012).

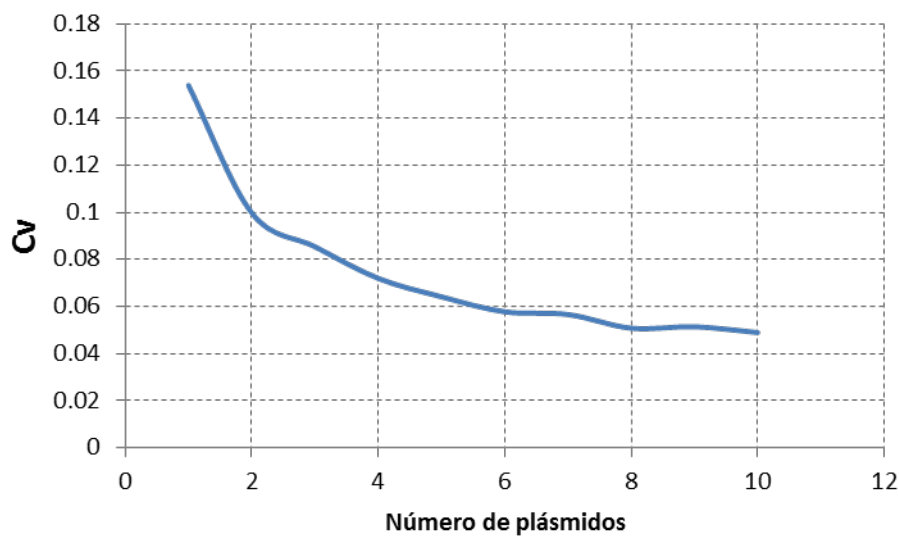


Figura 7. Efecto del número de copias de los plásmidos portadores de la compuerta lógica AND. Se grafica el coeficiente de variación del nivel de expresión de la GFP como función del número de copias del gen que la codifica.

Por otra parte, también se ha considerado el ruido molecular. El ruido es un proceso caracterizado y descrito en muchas áreas del conocimiento, se define como un proceso estocástico que produce variabilidad (Browsher y Swain, 2012). Biológicamente se le asocia el nombre de ruido bioquímico, definiéndolo para el caso de expresión de genes, como un proceso estocástico que produce variabilidad en la expresión. Además, se descompone en sus componentes intrínseca ( $\eta_{int}$ ) y extrínseca ( $\eta_{ext}$ ), y estas dos componentes se conjugan en procesos aleatorios (Paulsson, 2004). Una de las características del algoritmo estocástico es que considera el ruido intrínseco, y

consecuentemente lleva a un mejor retrato de la naturaleza aleatoria inherente a cualquier sistema genético de expresión. El ruido intrínseco nace de procesos como la unión de la RNA polimerasa al promotor o la eficiencia de la degradación, provocando que exista variabilidad entre organismos genéticamente idénticos (Raser y O'Shea).

La generación de variación es un problema en el sistema de compuertas biológicas ya que podría generar error en las señales de salida (Mano, 1994). Se observó que el ruido intrínseco simulado por el algoritmo estocástico provocó un nivel de variabilidad mayor al que se presenta en los resultados experimentales de Bonnet *et al.* (2013). El efecto negativo derivado del ruido intrínseco es que se complica la digitalización de la señal de salida de los dispositivos lógicos.

El problema de la alta variación puede ser enfrentado al considerar el número de copias de los genes que codifica al gen de la GFP. Experimentalmente, Bonnet y colaboradores clonaron el gen de la GFP en vectores originados del esqueleto del plásmido pSB4A5, que contiene el origen de replicación pSC101. Este origen de replicación produce un bajo número de plásmidos por célula, con un rango de entre 5 y 10 copias (Peterson y Philips, 2008). Igual que las pruebas *in vitro*, la modificación de la dosis del gen en los modelos afecta la variación de la señal de salida. Con el objetivo de analizar la variación de la señal de salida se utilizó el parámetro del coeficiente de variación (CV), que es el cociente de la desviación estándar y la media, y con ello fue posible medir el ruido de señal de salida tal y como se sugiere en la literatura (Swain *et al.*, 2002). Como resultado, en la Figura 7 se advierte que el aumento del número de copias disminuye la variabilidad de la expresión de la Gfp en la compuerta AND. Por otro lado, la Figura 9 permite apreciar que el aumento del número de copias plasmídico disminuye la variabilidad de expresión en todas las compuertas, y además, conduce a la homogenización de los niveles máximos de expresión entre los dispositivos. Otros estudios, tanto experimentales como teóricos, encuentran la misma tendencia e indican que el decremento en el número total de moléculas en un sistema de reacciones químicas se acompaña por un aumento en la amplitud de las fluctuaciones alrededor del punto de equilibrio anticipado por su evolución determinista en el tiempo, y viceversa (Paulsson *et al.*, 2004; Thattai y van Oudenaarden, 2001; y Rao *et al.*, 2002). Los resultados indican que esta característica se conserva para sistemas sin ningún tipo de retroalimentación, como el caso de las compuertas lógicas aquí expuestas. Así mismo, se demostró que si los parámetros cinéticos (y por lo tanto sus propensiones correspondientes) son alterados deliberadamente, se obtiene la misma disminución en el coeficiente

de variación cuando dichas alteraciones paramétricas provocan un aumento en el número de moléculas. Por lo tanto, ya sea por el número de plásmidos o por la modificación de los parámetros, la observación es que el número de moléculas de GFP como señal de salida tiene una relación inversa con la variabilidad de la señal. La consecuencia funcional es que uno podría mejorar el rango dinámico de los dispositivos elevando el número de moléculas de la señal de salida.

También es necesario mencionar que se advierte un efecto práctico considerable al momento de utilizar el modelado estocástico, y es que el tiempo de simulación aumenta con el número de variables y de moléculas. Esto ha sido advertido anteriormente, y se fundamenta en que aunque los algoritmos estocásticos generan realizaciones exactas de la ecuación química maestra, existe un precio computacional debido a que la aproximación estocástica simula cada reacción del sistema, y se vuelve extremadamente lenta cuando el número de eventos y moléculas es grande, cosa que también sucede con los sistemas químicos reales (Szallasi, Stelling y Periwal, 2006).



Compuerta	Tabla de verdad			=					=				
	Out1	Out2	Out	0	1	2	3	4	5	6	7	8	
AND	0	0	0										
	1	0	0										
	0	1	0										
	1	1	1	█									
NAND	0	0	1	█									
	1	0	1	█									
	0	1	1	█									
	1	1	0										
OR	0	0	0										
	1	0	1	█									
	0	1	1	█									
	1	1	1	█									
NOR	0	0	1	█									
	1	0	0										
	0	1	0										
	1	1	0										
XOR	0	0	0										
	1	0	1	█									
	0	1	1	█									
	1	1	0										
XNOR	0	0	1	█									
	1	0	0										
	0	1	0										
	1	1	1	█									

Figura 8. Desempeño predictivo de los dispositivos lógicos. En la primera columna se indica la función lógica, posteriormente la tabla de verdad correspondiente y los niveles de las señales de salida para cada estado de entrada. Las barras verde claro muestran los resultados normalizados de la señal de salida del modelo determinista y, en verde oscuro, los del modelo estocástico.

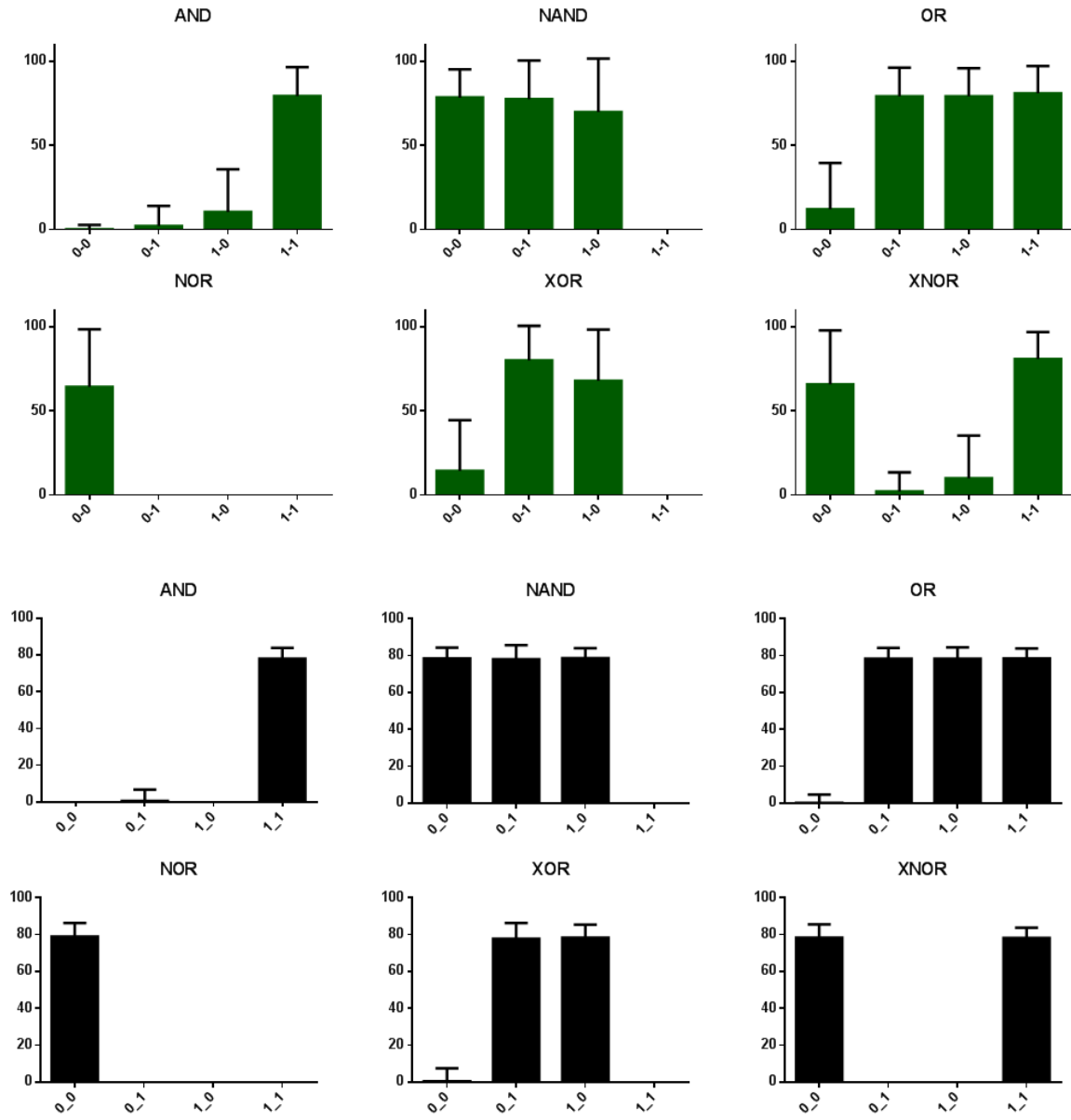


Figura 9. Efecto del número de plásmidos en todas las compuertas. En verde se presentan los niveles de expresión normalizados de las señales de salida arrojados por todos los dispositivos lógicos cuando el número de plásmidos portadores de las compuertas es 1. En negro muestra lo mismo para el caso en que el número de plásmidos por célula es de 7.

## ❖ Análisis de los circuitos extendidos

Utilizando las compuertas biológicas como bloques de construcción, de manera similar a como se realizan los circuitos en electrónica, se ensamblaron tres circuitos virtuales: un semi sumador, un sumador complejo y un decodificador 2x4. Estas construcciones están inspiradas en circuitos integrados, con lo que se pretende facilitar la introducción de los conceptos procedentes de la ingeniería electrónica y diseño a la biología de sistemas. El objetivo principal fue evaluar la aptitud de la aproximación *in vitro* para la construcción de circuitos de mayor complejidad y determinar las posibles características de los dispositivos extendidos no observables en las compuertas individuales.

El primer circuito es el semisumador y es capaz de realizar la suma aritmética binaria de dos únicos bits o señales de entrada, proporcionando a su salida un bit resultado de suma S y un bit de acarreo C. De tal manera que se construye acoplando las dos señales de entrada a una compuerta AND y a una compuerta OR en paralelo para dar lugar a dos señales de salida (Mano, 1994). En el panel A de la Figura 10 se bosqueja el circuito del semi sumador, mientras que en el panel B se expone la tabla de verdad.

El sumador completo, por otro lado, se obtiene conectando dos semi sumadores en paralelo, mejorando el proceso al cual se le añade una señal de entrada de acarreo. Con ello se pueden realizar sumas de más de un bit para las cuales se utiliza el acarreo de salida del bit anterior en el acarreo de entrada del siguiente, completando la suma correctamente (Mano, 1994). Este circuito consta de tres señales de entrada y dos de salida, y se construye utilizando 5 compuertas lógicas: un par de compuertas XOR, un par de compuertas AND y una compuerta OR. La configuración del sumador completo se esquematiza en el panel A de la Figura 11, y su correspondiente tabla de verdad en el panel B. Dada su estructura, su implementación experimental en el contexto biológico de Bonnet y colaboradores requiere 6 integrasas ortogonalmente funcionales.

Por último está el decodificador 2x4. Se trata de un dispositivo que permite traducir información binaria de una forma de codificación a otra. Consta de cuatro compuertas AND y dos inversores. Funciona con dos señales de entrada y cuatro de salida, pero para cada una de las combinaciones de entrada existe un solo canal de salida, es decir, sólo una de las señales de salida toma un valor alto (Tokheim, 1986). El procesamiento del decodificador se puede apreciar con mayor facilidad a

través de la tabla de verdad expuesta en el panel B de la Figura 13; en el panel A se muestra la arquitectura del decodificador.

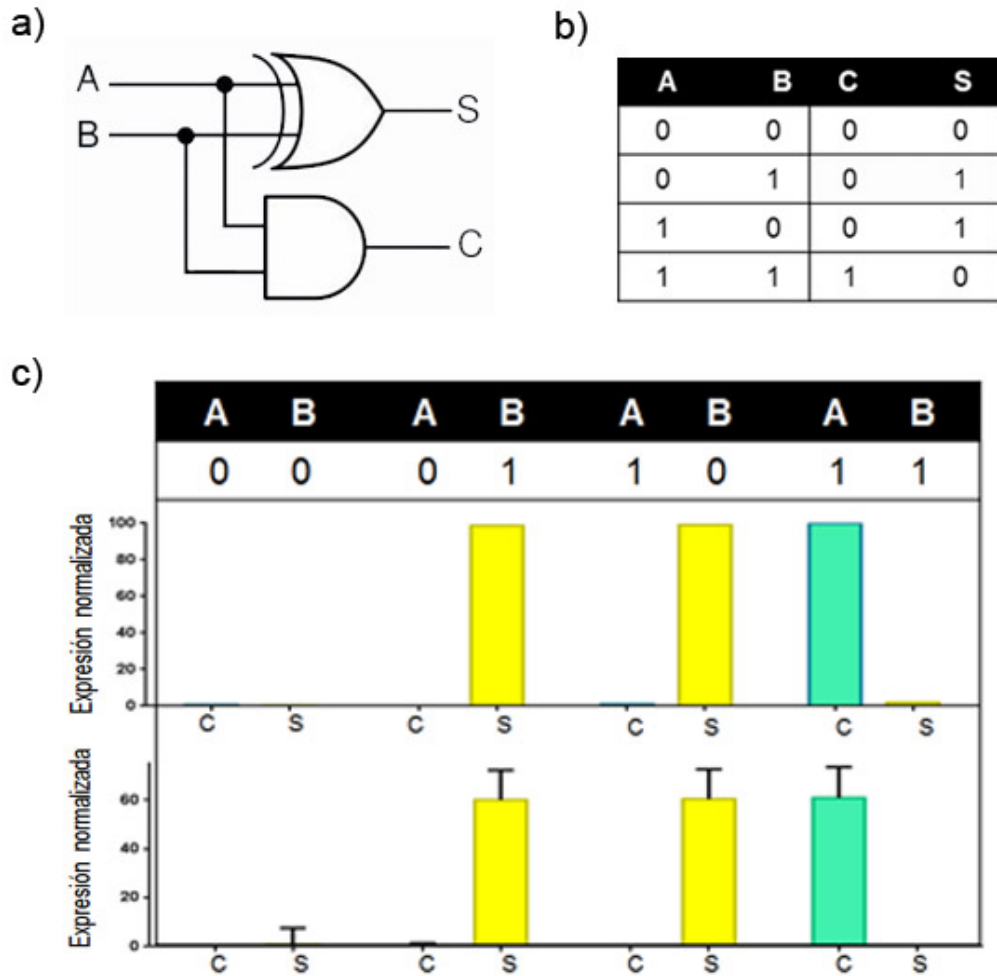


Figura 10. Semi sumador. En a) se tiene la estructura del circuito del semi sumador. En b) se presenta la tabla de verdad correspondiente. En c) se grafica la respuesta normalizada del circuito para cada estado de las señales de entrada a través de los modelos determinista (barras superiores) y estocástico (barras inferiores). Las barras celestes son para la salida del acarreador (C), y las amarillas para la suma (S).

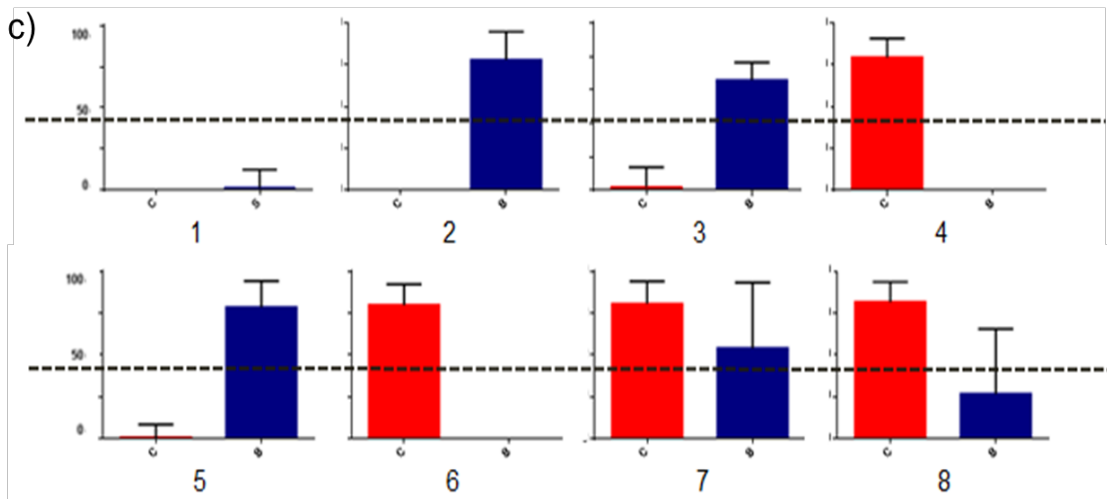
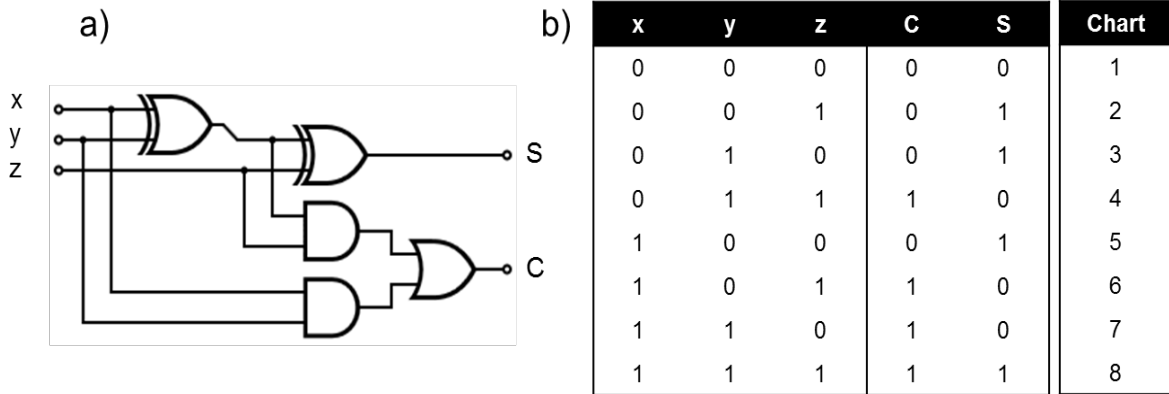


Figura 11. Sumador completo. En a) se muestra la arquitectura del sumador completo. En b) se explica la tabla de verdad del circuito. En c) están las predicciones del modelo estocástico del sumador; las barras rojas y azules representan los el nivel promedio de expresión en las señales de salida del acarreador (C) y de suma (S), respectivamente.

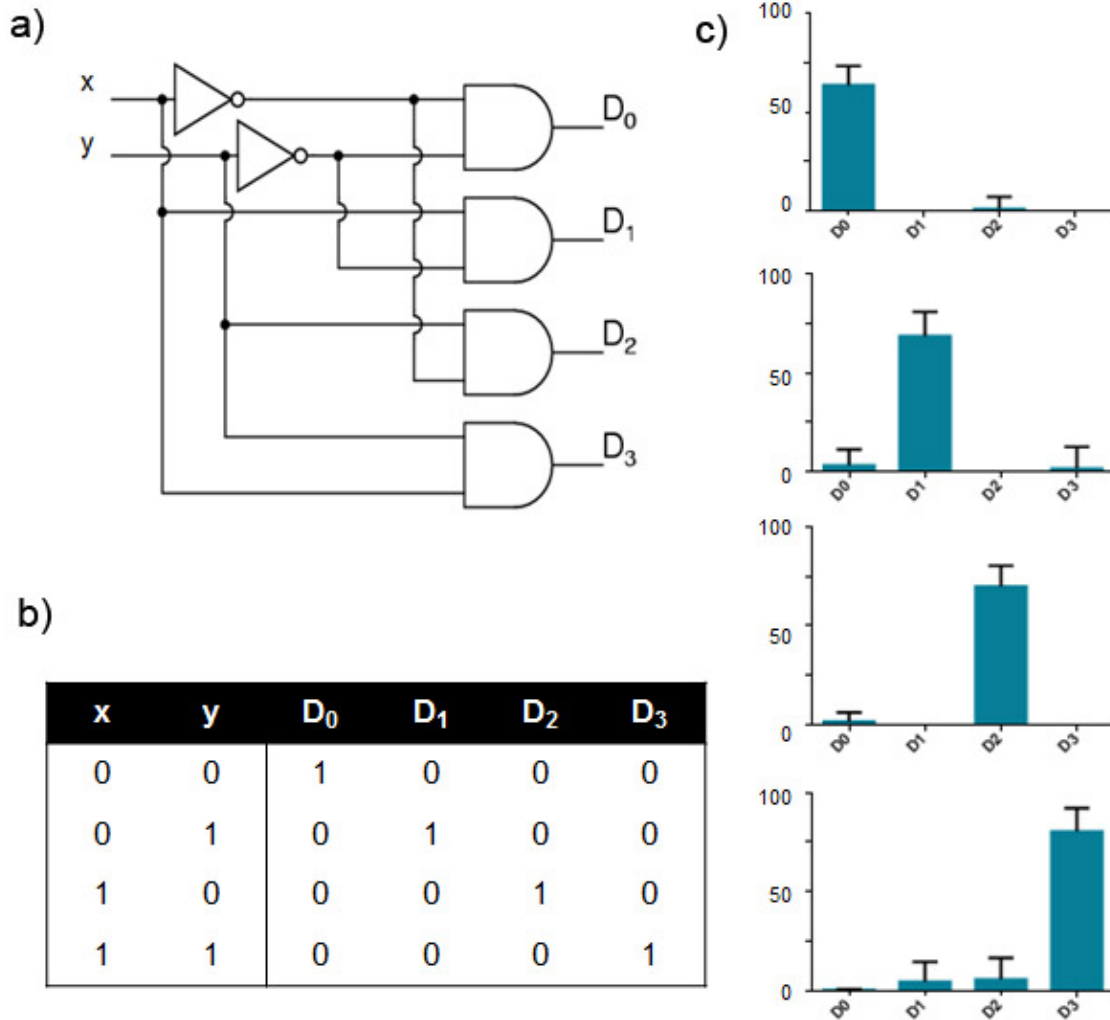


Figura 12. Decodificador 2x4. En a) se esquematiza la arquitectura del decodificador. En b) se encuentra la tabla de verdad del dispositivo. En c) se muestran las predicciones de la señal de salida del modelo estocástico para cada uno de los estados de las señales de entrada en orden descendente tal y como aparecen en la tabla de verdad.

La conexión entre compuertas lógicas implica que la señal de salida de un dispositivo toma el papel de señal de entrada en el siguiente dentro de la red. Según el esquema del transcriptor, los estados de las compuertas son controlados a través de la acción de las integrasas, de manera que son éstas enzimas quienes cumplen la función análoga a las señales de entrada. Por lo tanto, las señales de salida deben estar constituidas por los niveles de expresión de diferentes integrasas ortogonales para lograr el acoplamiento entre compuertas. Sin embargo, para el transcriptor sólo se realizó la

caracterización experimental de un par de integrasas, Bxb1 y TP901-1. Aunque existen integrasas alternativas aisladas de fagos y pertenecientes a la familia de las serín integrasas, no se cuenta con los parámetros de la dinámica de su actividad enzimática. Debido a ello, la construcción de los circuitos se realizó suponiendo que los parámetros para enzimas alternas son similares a los de Bxb1 y TP901-1, lo que probablemente sea correcto ya que la familia de las serín integrasas es bastante homogénea en tamaño y estructura (Groth y Carlos, 2004).

El primer circuito es el semi sumador. De su arquitectura puede concluirse que no hace falta más que un par de integrasas para su construcción, ya que las dos señales de entrada son comunes a las dos compuertas lógicas que lo constituyen. Su implementación es sencilla pero su funcionalidad es restringida, en comparación por ejemplo, con el sumador completo (Tokheim, 1986). Se evaluó la capacidad de la aproximación *in silico* para reproducir la tabla de verdad del semi sumador a través de los modelos determinista y estocástico; los resultados se encuentran resumidos en el panel C de la Figura 10. Las barras amarillas representan el nivel de expresión normalizado para la señal de salida de suma (S), mientras que las azules indican lo mismo para la salida del acarreador (C). Se puede observar que tanto el modelo determinista como el algoritmo estocástico procesan de forma acertada la información proveniente de las señales de entrada. Asimismo, el rango dinámico alcanzado con ambas aproximaciones es lo suficientemente alto como para establecer una digitalización del dispositivo.

El segundo dispositivo es el sumador completo. Como se describió anteriormente, es posible ensamblar un sumador completo a partir de dos semi sumadores y extender la capacidad de suma a más de un bit. Sin embargo también se debe notar que su construcción es más compleja pues se requieren de seis integrasas diferentes funcionales sin interferencia. Este dispositivo posee tres señales de entrada y puede aceptar ocho combinaciones distintas. Se construyó el circuito virtual a través de los modelos propuestos para estudiar su desempeño. En el panel C de la Figura 11 están los resultados obtenidos a través del modelo estocástico; las barras rojas señalan el nivel de expresión normalizado de la salida correspondiente al acarreador (C) y análogamente las barras azules lo hacen para la salida de la suma (S). Se observa que el rango dinámico del dispositivo permite distinguir los estados encendido y apagado en las señales de salida para seis de las combinaciones de entrada, sin embargo no ocurre lo mismo para las últimas dos combinaciones

(110) y (111). La digitalización del dispositivo falló por un problema inherente a las características biológicas del sistema y se revela claramente a través de la Figura 12, donde se muestran cuatro gráficas. Las gráficas encerradas en amarillo dan el seguimiento de la expresión de la integrasa resultante de la señal de salida de la primera compuerta XOR en el circuito (véase el panel A de la Figura 11) y de la señal de salida correspondiente a la suma (S), cuando la combinación de entrada es (110). Según esta combinación de entrada, la salida de la primer compuerta XOR debería ser baja o en estado apagado, y se ve que así ocurre para la mayor parte del tiempo de simulación. Sin embargo, existe una expresión transitoria durante los primeros 100 minutos dando lugar a un número máximo de aproximadamente 30 moléculas de integrasa. Dado que las integrasas tienen una gran especificidad por sus sitios de reconocimiento (Keravala *et al.*, 2006), el modelo predice que unas cuantas moléculas son suficientes para realizar el proceso de reconocimiento e inversión del terminador transcripcional en la segunda compuerta XOR del sumador. Como consecuencia de la inversión se permite entonces el paso de la ARN polimerasa a través del ADN generando una señal alta en la salida de suma (S) para algunas de las células simuladas. Por otro lado, las gráficas encerradas en verde muestran las trayectorias estocásticas para la misma integrasa y la señal de salida cuando la entrada es (011). En este caso la señal de salida (S) también debe ser baja o apagada y se observó que la respuesta del dispositivo fue correcta, a diferencia del caso anterior. El contraste ocurre porque con el estado de entrada (011), la respuesta de la primera compuerta XOR es alta, de manera que no hay ninguna expresión transitoria que afecte el procesamiento de las señales de entrada.

El último ensamblaje es el decodificador 2x4. Es importante mencionar que a diferencia de los circuitos anteriores, y debido a que cada combinación de las entradas genera una señal de encendido sólo en una de las vías de salida, el decodificador tiene aplicaciones biológicas más directas. Por ejemplo, Fan *et al.* (2012) proponen dispositivos lógicos con respuestas similares a las del decodificador hechos de nanoestructuras tetraédricas de ADN y los utiliza como sensores de iones, pequeñas moléculas, y cadenas de ADN complementario. Asimismo, Daringer y colaboradores (2014) utilizan biosensores de múltiples vías en superficie celular programados para registrar factores externos y modificar la fisiología celular con aplicaciones terapéuticas. Por lo tanto, el decodificador podría programarse para encender rutas celulares particulares con finalidades específicas.



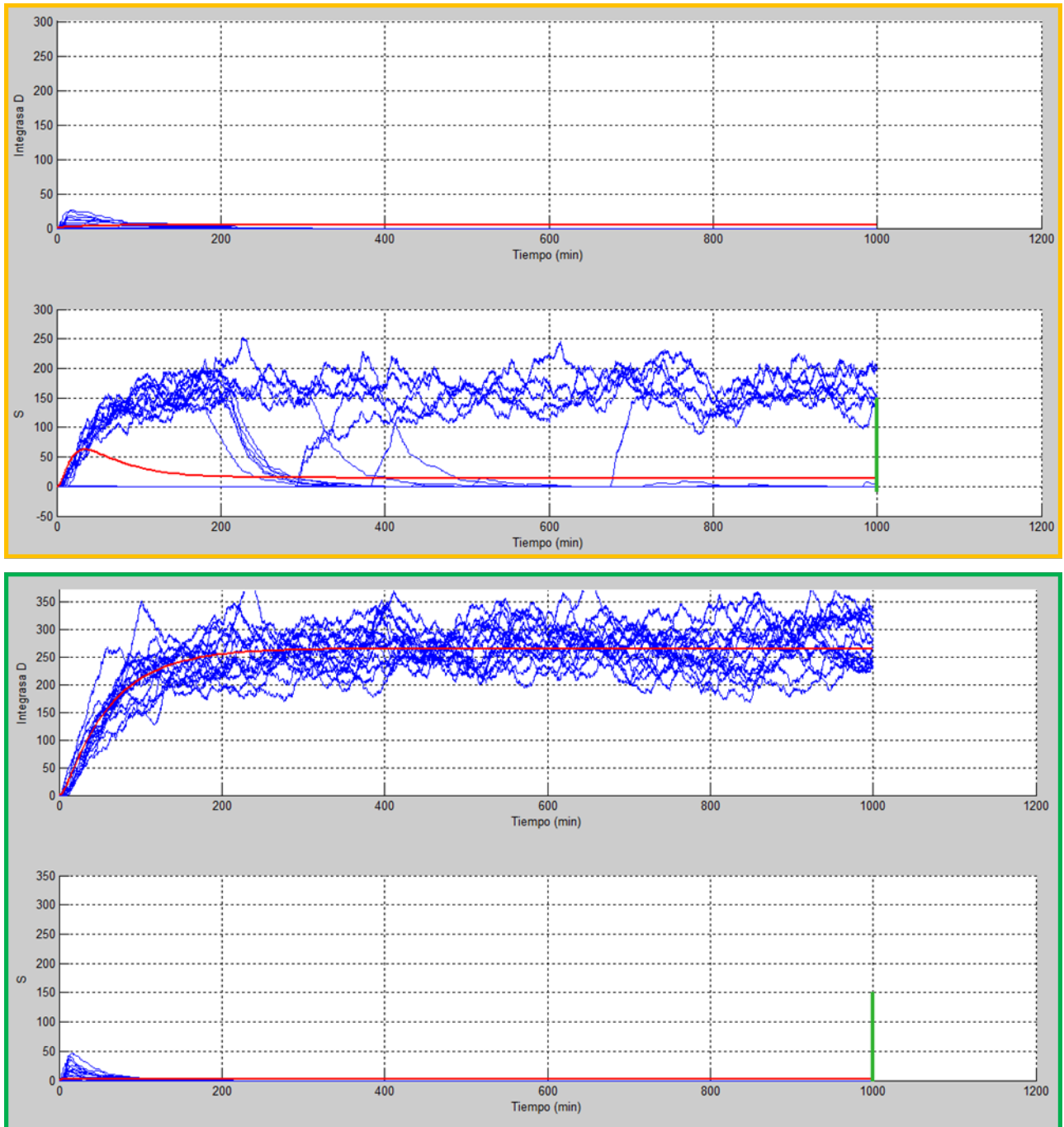


Figura 13. Expresión transitoria indeseada del sumador completo. Se muestra la expresión de la integrasa que constituye la señal de salida de la primera compuerta XOR del sumador completo, y la expresión de la proteína reportera en la salida de suma (S) del circuito. En amarillo se encierran las gráficas en que el estado de las señales de entrada es (110); y en verde cuando la entrada es (011). Se calcularon 30 trayectorias en cada caso.

Con la practicidad del decodificador en mente, y de la misma manera en que se hizo con los dos circuitos anteriores, se procedió a evaluar su funcionamiento. Para su construcción *in silico* fue necesario idear la arquitectura del dispositivo inversor ya que Bonnet y colaboradores (2013) no reportan esta función lógica, y consistió en un terminador transcripcional en su orientación inactiva, flanqueado por los sitios attB y attP de cualquiera de las integrasas. Los resultados del modelo estocástico predijeron problemas similares a los que se presentaron en la construcción del sumador completo y que son provocados por los transitorios. Para eliminar la expresión transitoria de las integrasas se propuso reducir el número de enzimas totales en su punto de equilibrio. Se observó que si se aumentan las tasas de degradación de las enzimas en tres órdenes de magnitud entonces el circuito reproduce de manera adecuada la tabla de verdad del decodificador (ver el panel C de la Figura 13). Lo que sucede al alterar la tasa de degradación y por lo tanto al disminuir el número de integrasas, es que hay una menor probabilidad de que la enzima realice el proceso de inversión de sus sitios de reconocimiento durante la expresión transitoria.

Experimentalmente, se podría lograr lo anterior disminuyendo la vida media de las integrasas a través de etiquetas de degradación. Una de ellas es la señal LVA, que consiste en una secuencia corta de 11 aminoácidos y que se une al extremo C-terminal de las proteínas; este péptido vuelve susceptible a la proteína a la que está unida a la acción de proteasas específicas. En ensayos con GFP se ha reportado que el uso de la etiqueta de degradación LVA disminuye la vida media de la proteína un 97% y en dos órdenes de magnitud (Andersen *et al.*, 1988). Por otro lado, Atkins y Raymind (2010) advierten que se puede acelerar la degradación del ARNm eliminando una porción del extremo 3' de la molécula, lo que la deja expuesta a la actividad de las exorribonucleasas celulares. Siendo que el número de copias de mensajeros de una especie es cercano o menor a 10, una rápida degradación de éstos tendría un gran impacto en el número de copias de la proteína correspondiente. Por lo tanto, si se combinan ambas estrategias es probable que se logre la disminución necesaria de la expresión de las integrasas y se podría alcanzar el comportamiento correcto del decodificador.

Se han reportado dificultades similares en otras construcciones donde también se trabaja con la conexión de dispositivos biológicos (Harold *et al.*, 2000). Estos estudios indican que, debido a la falta de sincronización en la progresión de las señales de entrada entre dispositivos, se generan errores transitorios en las señales de salida. Dicho de otra manera, si una señal de entrada alcanza al dispositivo correspondiente más rápido que las otras, se generarán respuestas incorrectas en las

señales de salida. Este fenómeno cobra mayor importancia a medida que se incrementa el tamaño y número de los circuitos puesto que podría propagar el efecto en órdenes de magnitud por encima de su origen (Cookson *et al.*, 2011). El concepto de asincronía en las compuertas es interesante y cabe la posibilidad de que, a través la adición de parámetros de retardo dependientes de las constantes de difusión de cada molécula dentro de los modelos propuestos, sea posible predecir este fenómeno.

Por el análisis de los circuitos anteriores se puede suponer que la caracterización de las compuertas lógicas como dispositivos individuales es insuficiente para predecir correctamente el desempeño de un dispositivo más complejo. Esta conjetura resulta de llevar la conceptualización de los circuitos a un contexto genético, donde se abre la probabilidad a efectos como el de la expresión transitoria indeseada de las integrasas, o bien, interferencia entre circuitos o un impacto en los recursos celulares, como ha sido referenciado en otros estudios (Arkin, 2008).

#### ❖ **Factibilidad en la implementación experimental de los circuitos**

Al igual que las características informáticas de cada dispositivo, también se debe considerar la contraparte experimental de los elementos biológicos con los que se podría realizar la construcción *in vitro*. A continuación se discutirán algunos puntos importantes.

**Ensamblaje físico de los circuitos.** Un aspecto esencial es el andamiaje físico de los circuitos y la capacidad del hospedero para almacenarlo. En este sentido, si se quiere implantar el decodificador 2x4 en *E. coli* como chasis, una estrategia consiste en introducir el programa en un par de plásmidos. El primero llevaría los genes codificantes para las integrasas y los promotores que controlan su expresión, además de un gen de resistencia como selector de bacterias transformantes, lo que daría un total aproximado de 6,860 pb de longitud. En el segundo plásmido estarían situadas las secuencias de dos compuertas inversoras, cuatro compuertas AND, cuatro promotores constitutivos, cuatro genes de proteínas reporteras y un gen de resistencia para selección, todo codificado en aproximadamente 6320 pb. Sin embargo, considerando el esqueleto de los plásmidos utilizados en experimentos de biología sintética (Shetty, Endy y Knight, 2008) que incluyen los

orígenes de replicación y secuencias intergénicas, la longitud total de los vectores se estima cerca de los 9,680 pb y 9,013 pb, respectivamente. Respecto a esto, se ha visto que es posible transformar cepas comunes de *E. coli* con vectores del rango de tamaño calculado (Sheng, Mancino y Birren, 1995). Incluso, en una construcción de dispositivos sintéticos se insertó un programa de 11 proteínas regulatorias y 38 partes genéticas adicionales, todo codificado en 21 kpb distribuidos en tres plásmidos (Loew *et al.*, 2010), por lo que, bajo la luz de estas consideraciones, debería ser posible la implementación de los circuitos propuestos.

**Aislamiento de las señales moleculares.** Otro aspecto importante que se mencionó anteriormente es la necesidad de múltiples integrasas con funcionalidad ortogonal. El hecho es que en electrónica, los circuitos integrados se fabrican conectando dispositivos a través de material conductor eléctricamente aislado, de manera que se evita la posibilidad de interferencia entre señales. Sin embargo, la célula carece de dicha capacidad de aislamiento, y las señales se transmiten a través de moléculas que pueden interactuar entre sí por las mismas leyes probabilísticas que se explicaron en el apartado del modelo estocástico. Por lo tanto, cada señal debe estar representada por una molécula diferente y deben ser recibidas por partes biológicas con alta especificidad. En el caso de las compuertas biológicas, las señales son acarreadas por las integrasas y recibidas por sus sitios de recombinación en el ADN. Hasta la fecha se ha realizado la caracterización molecular de distintas enzimas pertenecientes a la familia de las serín integrasas que reconocen sitios de recombinación específicas y con tamaños y estructuras similares (Zhang *et al.*, 2008; Chavez *et al.*, 2010; Keravala *et al.*, 2006). En la Tabla 6 se resume una búsqueda de dichas enzimas.

**Canales múltiples de entrada.** De igual modo, se requieren tantos promotores inducibles diferentes como señales de entrada tenga un circuito en particular. Este punto es relativamente más sencillo de cubrir ya que existe toda una colección de promotores inducibles probados en bacterias que responden a distintas moléculas inductoras e incluso existen asociaciones que se encargan de registrar partes biológicas para la construcción de circuitos sintéticos (Alterovitz, Muso y Ramoni, 2009).

Bacteriófago	Longitud de integrasa (aa's)	Hospedero
TP901-1	485	<i>Lactococcus lactis</i>
Bxb1	500	<i>Mycobacterium smegmatis</i>
R4	469	<i>Streptomyces parvulus</i>
φC31	605	<i>Streptomyces</i>
A118	452	<i>Listeria monocytogenes</i>
U153	452	<i>Listeria monocytogenes</i>
φFC1	465	<i>Enterococcus faecalis</i>
φRV1	469	<i>Mycobacterium tuberculosis</i>

Tabla 6. Ezimas pertenecientes a la familia de las serín integrasas.

Muchas construcciones sintéticas se obtienen utilizando estas partes para el control de la expresión genética y ejemplos de ello, por mencionar algunos, son los promotores inducibles por IPTG, por salicilato, por AHL, por propionato, y promotores sensibles a alta temperatura o alta osmolaridad (Brown, 2007). Sería necesario, además, generar modelos que describan la dinámica de apertura de estos promotores en función de la presencia o concentración de sus respectivos inductores con el propósito de representarlos como canales de entrada para su adición a los ensamblajes *in silico*.

**Consideraciones adicionales.** Existen otros factores que pueden influenciar la elección de algunas partes biológicas o estrategias de clonación sobre otras: las eficiencias de inversión de las integrasas podrían variar dependiendo de la longitud del ADN entre los sitios de recombinación (Ringrose *et al.*, 1999); la influencia de elementos regulatorios fuera de los dispositivos lógicos puede afectar la expresión en las señales de salida; la inactivación direccional de los elementos

regulatorios de las compuertas podría tener fugas (promotores y terminadores transcripcionales); e incluso los sitios de recombinación de las integrasas podrían afectar de alguna manera a los elementos de las compuertas lógicas (Siuti, Yazbek y Lu, 2013). En la medida en que estos cuestionamientos sean resueltos será posible realizar modelos *in vitro* más realistas y con capacidades predictivas más certeras, de manera que su realización permita a su vez construir circuitos biológicos *in vivo* con las características funcionales con las que fueron concebidos.

# CONCLUSIONES

---

Una de las direcciones del desarrollo de la biología sintética es el diseño rápido y preciso de circuitos biológicos que exhiban comportamientos programados previamente por el investigador. Con este fin, es necesario aplicar técnicas teóricas y experimentales para el diseño y caracterización de los bloques de construcción biológicos. En este contexto, los modelos *in vitro* no sólo permiten obtener una caracterización detallada de las partes biológicas, sino que también posibilitan la evaluación de sus modificaciones estructurales y la realización de predicciones que en última instancia reducen el número de experimentos. En este trabajo se investigó en detalle el funcionamiento de la arquitectura de seis compuertas lógicas y se propusieron tres circuitos virtuales ensamblados a partir de ellas. Se construyeron modelos teóricos tanto para las compuertas como para los circuitos y a partir de los resultados fue posible razonar que:

- La consideración de la naturaleza estocástica de las reacciones bioquímicas, manifestada como ruido celular intrínseco, es crucial para una representación más exacta de la red de interacción de estos dispositivos.
- Un número molecular alto aumenta el rango dinámico de los dispositivos lógicos y facilita el proceso de digitalización de la señal.
- Existe la posibilidad de expresión transitoria de integrasas intermediarias entre la señal de entrada y la de salida de circuitos construidos a partir de múltiples compuertas conectadas.
- La expresión transitoria de integrasas es suficiente para generar estados incorrectos en compuertas subsecuentes dentro del circuito y por lo tanto el procesamiento equivocado de la señal de entrada.

En resumen, a partir de este estudio se puede resolver que la biología sintética es beneficiada por los métodos computacionales, y que el proceso iterativo del diseño de modelos, la validación con datos experimentales y su actualización acelera el desarrollo de este campo.



# APÉNDICE A. Descripción del planteamiento estocástico.

---

Describiendo en términos probabilísticos, la base del método de Gillespie está definida por  $p(\tau, j \mid x, t)d\tau$ . Esta expresión denota la probabilidad de que, dado un estado del sistema determinado  $X(t)=x$ , la próxima reacción en el sistema ocurra en el intervalo de tiempo infinitesimal  $[t + \tau, t + \tau + d\tau)$ , y que la reacción elegida sea  $R_j$ . Formalmente, la expresión relaciona la función de densidad de probabilidad de dos variables aleatorias: el tiempo que transcurre entre cada reacción ( $\tau$ ) y el índice que identifica a la reacción seleccionada ( $j$ ).

Considerando las leyes de probabilidad y llevando al límite  $d\tau \rightarrow 0$  se obtiene una ecuación diferencial cuya solución resulta en  $P_0(\tau \mid x, t) = \exp(-a_0(x) \tau)$ , donde  $P_0(\tau \mid x, t)$  es la probabilidad de que ninguna reacción ocurra en el intervalo de tiempo  $[t, t + \tau)$  y  $a_0(x)$  está dada por

$$a_0 = \sum_{j=1}^M a_j(x) \tag{5}$$

Internando este resultado en la ecuación de  $p$ , tenemos

$$p(\tau, j \mid x, t) = a_j(x) \exp(-a_0(x) \tau) \tag{6}$$

La ecuación anterior es el fundamento matemático para la aproximación de la simulación estocástica. Implica que la relación de las funciones de densidad de  $\tau$  y  $j$  se pueden expresar como el producto de la función de densidad  $\tau$ ,  $a_0(x) \exp(-a_0(x) \tau)$ , y de la función de densidad de  $j$ ,  $a_j(x)/a_0(x)$ .

Con esta información es posible implementar el algoritmo de Gillespie de la manera en que se describe en el apartado de Materiales y Metodología. En el proceso se generan dos números aleatorios  $r_1$  y  $r_2$  desde una distribución uniforme y se seleccionan  $\tau$  y  $j$  de acuerdo a

$$\tau = \frac{1}{a_0(x)} \ln \frac{1}{r_1} \quad 7$$

$j$  = el número entero más pequeño que satisfaga la condición

$$\sum_{j=1}^j a_j(x) > r_2 a_0(x) \quad 8$$

# REFERENCIAS

---

- Alterovitz G, Muso T y Ramoni MF. 2009. The challenge of informatics in synthetic biology: from biomolecular networks to artificial organisms. *Briefings in Bioinformatics*: 2 (1), pp. 80-95.
- Anderson JC, Clarke EJ, Arkin AP y Voigt CA. 2006. Environmentally controlled invasion of cancer cells by engineered bacteria. *J. Mol. Biol.*: 355, 619-627.
- Arkin A. 2008. Setting the standard in synthetic biology. *Nature Biotechnol.*: 26, pp. 771–774.
- Arpino JA, *et al.* 2013. Tuning the dials of synthetic Biology. *Microbiology*: 159, pp. 1236-1253.
- Atkins JF y Raymind FG. 2010. *Recoding: expansion of decoding rules enriches gene expression*. Springer Science & Business Media, EUA.
- Barrett CL *et al.* 2006. Systems biology as a foundation for genome-scale synthetic biology. *Curr. Opin. Biotechnol.*: 17: pp. 488–492.
- Bashor CJ, Helman NC, Yan S y Lim WA. 2008. Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics. *Science*: 319, pp. 1539-1543.
- Basu S, Gerchman Y, Collins CH, Arnold FH y Weiss R. 2005. A synthetic multicellular system for programmed pattern formation. *Nature*: 434, pp. 1130–1134.
- Bernstein *et al.* 2002. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *PNAS*: 99 (15), pp. 9697-9702.
- Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J y Phillips R. 2005. Transcriptional regulation by the numbers: models. *Current opinion in genetics & development*: 15 (2), pp. 116-124.
- Bonnet J, Subsoontorn P y Endy D. 2012. Rewritable digital data storage in live cells via engineered control of recombination directionality. *PNAS*: 109 (23), pp. 8884-8889.
- Bonnet J, Yin P, Ortiz ME, Subsoontorn P y Endy D. 2013. Amplifying genetic logic gates. *Science*: 340 (599), pp. 599-603.
- Brown J. 2007. The iGEM competition: building with biology. *IETSynth. Biol.*: 1, pp. 3–6.
- Browsher CG y Swain PS. 2012. Identifying sources of variation and flow of information in biochemical networks. *PNAS*: 109 (20), pp. 1320–1328.

- Carroll P, Muttucumaru DG y Parish T. 2005. Use of a tetracycline-inducible system for conditional expression in *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*. *Appl. Environ. Microbiol.*: 71 (6), pp. 3077-3084.
- Chavez CL *et al.* 2010. Kinetics and longevity of  $\Phi$ c31 integrase in mouse liver and cultured cells. *Human Gene Therapy*: 21, pp. 1287-1297.
- Cheng AA y Lu TK. 2012. Synthetic biology: an emerging engineering discipline. *Annu. Rev. Biomed. Eng.*: 14, pp. 155-178.
- Chunhai Fan, *et al.* 2012. Reconfigurable three-dimensional DNA nanostructures for the construction of intracellular logic sensors. *Angewandte Chemie*: 51, pp. 9020–9024.
- Cookson NA, *et al.* Queueing up for enzymatic processing: correlated signaling through coupled degradation. *Mol. Syst. Biol.*: 7 (561).
- Daringer NM *et al.* 2014. Modular extracellular sensor architecture for engineering mammalian cell-based devices. *ACS Synth. Biol.*
- Elowitz MB y Leibler S. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature*: 403, pp. 335–338.
- Endy D. 2005. Foundations for engineering biology. *Nature*: 438, pp. 449-453.
- Fritz BR, Timmerman LE, Daringer NM, Leonard JN y Jewett MC. 2010. Biology by design: from top to bottom and back. *J. Biomed. Biotechnol.*: 232016.
- Gardner TS, Cantor CR y Collins JJ. 2000. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*: 403, pp. 339–342.
- Gillespie DT. 1992. *Markov processes: an introduction for physical scientists*. MA: Academic Press, EUA.
- Glass L, Kauffman SA. 1973. The logical analysis of continuous, non-linear biochemical control networks. *J Theor Biol.*: 39(1), pp. 103-29.
- Goñi-Moreno A y Amos M. 2012. A reconfigurable NAND/NOR genetic logic gate. *BMC Systems Biology*: 6 (126).
- Grigorova IL, Phleger NJ, Mutalik VK y Gross CA. 2006. Insights into transcriptional regulation and sigma competition from an equilibrium model of RNA polymerase binding to DNA. *PNAS*: 103, pp. 5332-5337.
- Groth AC y Carlos MP. 2004. Phage integrases: biology and applications. *J Mol Biol*: 335 (3), pp. 667-678.
- Guet CC, Elowitz MB, Hsing W y Leibler S. 2002. Combinatorial synthesis of genetic networks. *Science*: 296, pp. 1466–1470.

- Guzman LM, Belin D, Carson MJ y Beckwith J. 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose P<sub>BAD</sub> promoter. *Journal of Bacteriology*: 177 (14), pp.4121-4130.
- Harold A, *et al.* 2000. Amorphous computing. *Commun. ACM.*: 43, pp. 74–82.
- Hoffmann G, Ostermeir R, Müller H, Bitterlich N, Martus P y Neumaier M. 2008. SimChip-computer simulation of mRNA steady states. *Clin. Lab.*: 54, pp. 19-24.
- Hoffman-Sommer M, Supady A, y Klipp E. 2012. Cell-to-cell communication circuits: quantitative analysis of synthetic logic gates. *Frontiers in Physiology*: 3 (287).
- Huh D y Paulsson J. 2010. Random partitioning of molecules at cell division. *PNAS*: 108 (36), pp. 15004-15009.
- Ishihama Y, *et al.* 2008. Protein abundance of the *Escherichia coli* cytosol. *BMC Genomics*: 9 (102).
- Iyer V y Struhl K. 1996. Absolute mRNA levels and transcriptional initiation rates in *Saccharomyces cerevisiae*. *PNAS*: 93, pp. 5208–5212.
- J.B. Andersen *et al.* 1988. New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Applied and Environmental Microbiology*: 64, pp. 2240–2246.
- Keravala A *et al.* 2006. A diversity of serine phage integrases mediate site-specific recombination in mammalian cells. *Mol Gen Genomics*: 276 (135), pp. 135-146.
- Khalil AS y Collins JJ. 2010. Synthetic biology: applications come of age. *Nat. Rev. Genet.*: 11 (5), pp. 367-379.
- Kierzek AM, Zaim J y Zielenkiewicz P. 2001. The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression. *J. Biol. Chem.*: 276, pp. 8165–8172.
- Kim AI, Ghosh P, Aaron MA, Bibb LA, Jain S y Hatfull GF. 2003. Mycobacteriophage Bxb1 integrates into the *Mycobacterium smegmatis* groEL1 gene. *Mol. Microbiol.*: 50, pp. 463–473.
- Lim WA. 2010. Designing customized cell signalling circuits. *Nat. Rev. Mol. Cell Biol.*: 11, pp. 393–403.
- Loew R, Heinz N, Hampf M, Bujard H y Gossen M. 2010. Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnology*: 10 (81).
- Longo D y Hasty J. 2006. Dynamics of single-cell gene expression. *Molecular Systems Biology*: 1 (64).
- Mano MM. 1994. *Arquitectura de computadoras*. Pearson Educación, España.
- McAdams HH, Arkin A. 2000. Towards a circuit engineering discipline. *Curr. Biol.*: 10, pp.318.

- Monod J, Jacob F. 1961. General conclusions: telenomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb. Symp. Quant. Biol.*: 26, pp. 389–401.
- Moon TS, Lou C, Tamsir A, Stanton BC y Voigt CA. 2012. Genetic programs constructed from layered logic gates in single cells. *Nature*: 491, p. 249-253.
- Morris MK, Saez-Rodriguez J, Sorger PK y Lauffenburger DA. 2010. Logic-based models for the analysis of cell signaling networks. *Biochemistry* : 49, pp. 3216–3224.
- Paulsson J. 2004. Summing up the noise in gene networks. *Nature*: 427, pp. 415-418.
- Peterson J, Phillips GJ. 2008. New pSC101-derivative cloning vectors with elevated copy numbers. *Plasmid*: 59 (3), pp. 193-201.
- Pollock TJ y Nash HA. 1983. Knotting of DNA caused by a genetic rearrangement. Evidence for a nucleosome-like structure in site-specific recombination of bacteriophage lambda. *J. Mol. Biol.* : 170, pp. 1–18.
- Pratt JM, *et al.* 2002. Dynamics of protein turnover, a missing dimension in proteomics. *Mol. Cell Proteomics.*: 1, pp. 579–591.
- Ptashne M. 2004. A genetic switch. Cold Spring Harbor Lab Press, USA.
- Rao C, Wolf D y Arkin A. 2002. Control, exploitation and tolerance of intracellular noise. *Nature*: 420, pp. 231-237.
- Raser MJ y O’Shea KE. 2005. Noise in gene expression: origins, consequences, and control. *Science*: 309, pp. 2010-2013.
- Regot S *et al.* 2011. Distributed biological computation with multicellular engineered networks. *Nature*: 469, pp. 207-211.
- Reyes O, Gottesman M y Adhya S. 1979. Formation of lambda lysogens by IS2 recombination: gal operon-lambda pR promoter fusions. *Virology*: 94, pp. 400–408.
- Ringrose L, Chabanis S, Angrand PO, Woodroffe C y Stewart AF. 1999. Quantitative comparison of DNA looping *in vitro* and *in vivo*: chromatin increases effective DNA flexibility at short distances. *EMBO J.*: 18, pp. 6630–6641.
- Sauro HM. 2011. Enzyme kinetics for systems biology. Ambrosius Publishing, EUA.
- Savageau MA. 1974. Comparison of classical and autogenous systems of regulation in inducible operons. *Nature*: 252, pp. 546–549.
- Sayut DJ, Niu Y y Sun L. 2009. Construction and enhancement of a minimal genetic AND logic gate. *Appl. Environ. Microbiol.*: 75 (3), pp. 637-642.
- Scharle TW. 1965. Axiomatization of propositional calculus with Sheffer functors. *Notre Dame J. Formal Logic*: 6, pp. 209–217.

- Sheng Y, Mancino V y Birren B. 1995. Transformation of *Escherichia coli* with large DNA molecules by electroporation. *Nucleic Acids Research*: 23 (11), pp. 1990-1996.
- Shetty R, Endy D y Knight T. 2008. Engineering BioBrick vectors from BioBrick parts. *Journal of Biological Engineering*: 14, pp. 2:5.
- Shnerb NM, Louzoun Y, Bettelheim E y Solomon S. 2000. The importance of being discrete: life always wins on the surface. *PNAS*: 97 (19), pp. 10322-10324.
- Siuti P, Yazbek J y Lu TK. 2013. Synthetic circuits integrating logic and memory in living cells. *Nature Biotechnology*: 31 (5), pp. 448-452.
- Stamatakis M y Mantzaris NV. 2009. Comparison of deterministic and stochastic models of the lac operon genetic network. *Biophys. J.*: 97(9), pp. 2651.
- Stavreva DA, *et al.* 2009. Ultradian hormone stimulation induces glucocorticoid receptor-mediated pulses of gene transcription. *Nat. Cell. Biol.*: 11, pp. 1093–1102.
- Stoll SM, Ginsburg DS y Calos MP. 2002. Phage TP901-1 site-specific integrase functions in human cells. *J. Bacteriol.*: 184 (13): pp. 3657-3663. (w)
- Swain PS, Elowitz MB y Siggia ED. 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression. *PNAS*: 99 (20), pp. 12795-12800.
- Szallasi Z, Stelling J y Periwé V. 2006. System modeling in cellular biology, from concepts to nuts and bolts. MIT press, EUA.
- Tabor JJ, *et al.* 2009. A synthetic genetic edge detection program. *Cell*: 137, pp. 1272–1281.
- Tamsir A, Tabor JJ y Voigt CA. 2011. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature*: 469 (13), pp. 212-215.
- Taniguchi Y, *et al.* 2010. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science*: 329, pp. 533-538.
- Thattai M y Van Oudenaarden A. 2001. Intrinsic noise in gene regulatory networks. *PNAS*: 98 (8614).
- Tokheim RL. 1986. *Electrónica digital*. Editorial Reverté, España.
- Twycross J, Band LR, Bennett MJ, King JR y Krasnogor N. 2010. Stochastic and deterministic multiscale models for systems biology: an auxin-transport case study. *BMC Systems Biology*: 4 (34).
- Wang B, Kitney RI, Jolv N y Buck M. 2011. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature Communications*: 2 (508).
- Watson JD, Hopkins NH, Roberts JW, Steitz JA y Weiner AM. 2003. *Molecular Biology of the gene*. Benjamin Cummings, USA.

- Weber W, *et al.* 2007. A synthetic time-delay circuit in mammalian cells and mice. PNAS: 104, pp. 2643–2648.
- Widmaier DM, *et al.* 2009. Engineering the *Salmonella* type III secretion system to export spider silk monomers. Mol. Syst. Biol.: 5 (309).
- Zhang J y Tianshou Z. (2014). Promoter-mediated transcriptional dynamics. Biophysical Journal: 106, pp. 479-488.
- Zhang L, Ou Xijun, Zhao G y Ding X. 2008. Highly efficient in vitro site-specific recombination system based on Streptomyces phage  $\Phi$ BT1 integrase. J. Bacteriol: 190 (19), pp. 6392- 6397.