

# Investigando enfermedades complejas con modelación matemática de sistemas nanobiológicos

# a

Luis Adolfo Torres González\*

"Tus alegrías y tus penas,  
tus recuerdos y tus ambiciones,  
tu identidad y tu libre albedrío,  
no son sino el comportamiento  
de un vasto conglomerado  
de células nerviosas..."  
Francis Crick, 1994

## Introducción

El Proyecto del Genoma Humano (PGH) culminado en el año 2003 trajo consigo una revolución de alcances insospechados para el siglo XXI, tanto científica como tecnológicamente. Dicho proyecto consistió en determinar las posiciones relativas de todos los 3 000 millones de nucleótidos o pares de bases (código de la vida) e identificar los 25 000 a 30 000 genes (secuencias de nucleótidos) presentes en él. El código fue abierto y ahora es preciso investigarlo y «descifrarlo» minuciosamente para entender cómo funcionan los mecanismos moleculares en los humanos. Esto contribuirá, además de otros intereses, en la solución de problemas de salud de nuestro país, al saber de qué manera producir nuevas vacunas y

fármacos que respondan a la intervención de enfermedades, y sobre todo, al entender distintas enfermedades complejas, las cuales cada vez cobran mayor fuerza en todos los sectores sociales de nuestro país sin tener respuestas contundentes de solución hasta ahora, como son el cáncer y la diabetes.

La tarea no es fácil, ya que todo esto será imposible sin la ayuda de equipos de *investigación interdisciplinaria*, donde las matemáticas, las ciencias básicas (física, química y biología), la biología molecular, las ciencias computacionales y las ciencias biomédicas, unen esfuerzos para buscar y determinar soluciones a estos *sistemas biológicos complejos*, cuyos entendimientos se encuentran seguramente en los mecanismos que subyacen en el orden de nanoescalas ( $10^{-9}$  metros o 0.000000001 metros), de ahí el nombre de nanobiología.

En este siglo el reto se vuelve mayor, debido a que la intención no es analizar sistemas de pocos elementos, sino la integración de muchos elementos de

estudio interactuando hasta conformar sistemas biológicos (*systems biology*). Por ejemplo, en el siglo pasado se hablaba de "genética", es decir, se estudiaba un gen o algunos genes; sin embargo, ahora nos interesa analizar la red de miles de genes, por ello el nombre de «genómica» y también de ahí, la importancia de usar supercomputadoras para su análisis.

Nos enfrentamos a una integración de elementos en lugar de una *reducción*. Si hablamos de los constituyentes básicos del ser humano, entonces nos interesa saber sobre las interacciones entre "redes" de metabolitos, proteínas, genes, nucleótidos; es decir, nos interesa investigar la maravillosa maquinaria que nos constituye. Por consiguiente, el conocer el funcionamiento de estas redes permitirá formular herramientas para encontrar soluciones contundentes a diversas enfermedades.

Un esfuerzo de estas investigaciones es desarrollado en el Departamento de Ciencias Básicas (DCB) por la comunidad de Investigación en Ingeniería y Ciencias Aplicadas (IICA) de la Universidad Iberoamericana (UIA) León, con el proyecto denominado: "Clustering de genes utilizando el modelo de Potts y simulación Monte Carlo". Dicho proyecto es apoyado actualmente por el Consejo de Ciencia y Tecnología del Estado de Guanajuato (CONCYTEG). Este proyecto, en su primera fase, está orientado a generar herramientas bioinformáticas –algoritmos computacionales basados en modelos físicos y matemáticos– para analizar datos de expresión de genes (microarreglos) y así obtener una explicación de la regulación entre los distintos genes. Así mismo, esto llevará a la ubicación de módulos funcionales que indiquen como se regulan

ciertos grupos de genes, y entonces quizás, a conocer la causa del porqué algunos genes se activan (o inhiben) produciendo enfermedades en los humanos. De la misma manera, este proyecto abre las posibilidades de ser uno de los primeros grupos del país en conformar una investigación interdisciplinaria con incidencia directa en el sector salud, además de constituir en las instalaciones de la UIA-León, el primer laboratorio de bioinformática y análisis de datos con estas características en el estado de Guanajuato.

## Investigación interdisciplinaria

La investigación interdisciplinaria en México se encuentra en sus primeras fases de desarrollo; sin embargo, el proyecto "Clustering de genes utilizando el modelo de Potts y simulación Monte Carlo" es considerado como uno de los primeros en el país y en el mundo de tipo interdisciplinario, además la temática ha sido considerada por una revista internacional de arbitraje como una propuesta relevante en investigación y con proyección significativa en el siglo XXI (Ibarra-Junquera *et al.*, 2005)<sup>1</sup>.

La genómica y en general las tecnologías *omics* que fueron impulsadas a finales del siglo pasado, han fortalecido la investigación interdisciplinaria a través de una rama llamada bioinformática (Campbell y Heyer, 2003), la cual responde a herramientas para el análisis de millones de datos de genes, proteínas, enzimas y metabolitos que se han obtenido en los laboratorios biotecnológicos de todo el mundo.

Un antecedente histórico de investigaciones interdisciplinarias de gran relevancia mundial fue el desarrollado por el físico Erwin Schrödinger en su libro intitolado *What is life? The physical aspect of the living*

\*Coordinador de Física del  
Departamento de  
Ciencias Básicas, UIA León.  
Miembro del SNI.  
adolfo.torres@leon.uia.mx

<sup>1</sup> La revista *Virtual Journal of Biological Physics Research* ha considerado este artículo como de punta, en el ámbito internacional.

## La investigación interdisciplinaria será la clave en el siglo XXI para encontrar respuestas a múltiples enfermedades

cell, publicado por primera vez en 1944, el cual es considerado como el origen de la biología molecular. Schrödinger fue uno de los primeros en abordar temas entre física moderna y sistemas vivos cuestionándose, por ejemplo, si la vida o los sistemas vivos violan las leyes de la física. Por otro lado, el trabajo interdisciplinario del físico Francis Crick y del biólogo James Watson, propuso un modelo de doble hélice en 1953: el ADN o ácido desoxirribonucleico o «molécula de la vida», el cual se encuentra presente en las células y contiene toda la información de los 3 000 millones de nucleótidos y la información de los genes o unidades básicas de herencia. Ambos científicos recibieron el premio Nobel por su notable contribución.

En este artículo se presentan algunos tópicos generales sobre las investigaciones realizadas desde la IICA, en particular sobre el *clustering* o agrupación de genes. Cabe mencionar que dicho proyecto se trabaja como parte de una red de grupos de investigación y colaboradores, en entre ellos, el doctor Haret Rosu Barbus del Instituto Potosino de Investigación Científica y Tecnológica (IPICYT); el doctor José Luis González del Centro Universitario de los Lagos (CU-Lagos), de la Universidad de Guadalajara (UdeG), y el doctor Ulises Pérez, del Hospital Central Regional de León.

### Clustering de genes

Los modelos físicos y matemáticos son herramientas indispensables en el procesamiento de información biológica de la célula viva, desde una

vía de señalización simple hasta una red compleja regulatoria. La predicción de modelos detallados de grandes redes podrá revolucionar la forma en que los investigadores estudian las enfermedades complejas. En esta investigación se hace un esfuerzo para presentar dos diferentes escalas de redes de genes. Mostramos las ventajas de la simulación Monte Carlo y el análisis dinámico en el campo de las redes génicas. En este sentido, identificamos dos problemas principales en dos escalas diferentes de redes de genes: a) el proceso de la regulación en redes génicas usando el modelo de Potts y b) el comportamiento dinámico de pequeños circuitos de genes. Con respecto al primer inciso, analizamos el *clustering* de tales redes considerando los datos de expresión de microarreglos que definen las dimensiones en la red. Además, investigamos el *clustering* de tales redes aplicando el modelo espinorial de Potts a los datos de expresión de genes introduciendo una interacción entre puntos vecinos. En el caso de las pequeñas redes de genes, se usa un sensor de *software* para monitorear el desarrollo de esos procesos fundamentales en detalle. Es importante mencionar que una de las publicaciones logradas como antecedente de este proyecto de investigación fue incluida en la lista de los mejores tópicos (mundialmente) contemporáneos en el área de física biológica de la revista de investigación *Virtual Journal of Biological Physics Research*, en agosto de 2005 (L.A. Torres et al., 2007).

Como parte de esta investigación, se implementó un laboratorio básico de bionanoinformática y análisis de datos en la UIA León, el cual permitirá correr las simulaciones computacionales y tener conexión a distintos nodos de supercomputadoras en el país o en el

extranjero. Los datos que se utilizaron en las primeras pruebas de los algoritmos son provenientes de la base de datos del doctor Julio Collado, del Centro de Ciencias Genómicas, de la Universidad Nacional Autónoma de México (UNAM) Cuernavaca. Los datos son de expresión de genes (microarreglos) para distintas condiciones experimentales de la bacteria *E. Coli*. Cabe mencionar que la "expresión de un gen" es el proceso por el cual un gen se activa en una célula para elaborar ARN (ácido ribonucleico) y proteínas. La expresión de un gen puede medirse al observar el ARN o la proteína elaborada con el ARN, o bien, lo que la proteína hace en una célula (National Cancer Institute, 2008).

Los microarreglos de ADN —o chips de ADN— permiten medir simultáneamente el nivel de transcripción<sup>2</sup> para cada gen en un genoma (Sketel, 2003). Mediante un arreglo experimental de robots y lectores se extrae la información simultánea de la expresión de miles de genes del organismo. Los datos aparecen en términos de señales de fluorescencia, las cuales son obtenidas al incidir un haz de láser sobre las muestras de interés. Cada microarreglo es escaneado con un láser que produce para cada tinte un mapa digital o imagen con *intensidades de fluorescencia* para cada píxel (ver figura 1).

Una vez obtenida la selección de datos experimentales se procede a generar y construir modelos matemáticos que permitan entender la regulación de los genes, para ello es necesario elaborar un análisis estricto de los de datos y construir el *clustering* de distintos genes, de tal manera que se pueda obtener información de los nodos funcionales. En nuestro caso usamos la simulación Monte Carlo<sup>3</sup> y el

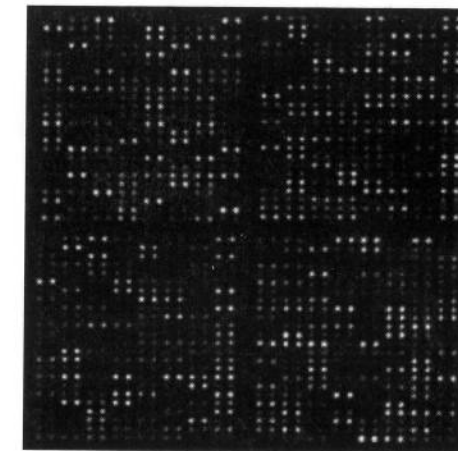


Figura 1 Microarreglo. Cada círculo representa información de la transcripción de cada gen. La intensidad y el color de cada círculo representan los niveles de expresión de cada gen.

modelo de Potts para modelar los estados de "encendido" o "apagado" (gen activado o inhibido) de cada gen en una red de miles de genes.

En una segunda fase del proyecto se obtendrán microarreglos de todo el perfil de genes en el humano, obteniendo así microarreglos de distintos tipos de cáncer y diabetes de personas en la ciudad de León y de distintos sitios del estado de Guanajuato con la finalidad de obtener un patrón genómico de una muestra representativa. De esta manera se investigan los mecanismos de regulación de los genes que *activan* o *inhiben* determinados tipos de cáncer o en general ciertas enfermedades crónicas en la población.

### Un ejemplo sencillo de modelación y simulación de redes de regulación génica

Para entender el funcionamiento de organismos a nivel molecular necesitamos conocer cuáles genes son expresados

<sup>2</sup> La transcripción del ADN es el primer proceso de la expresión genética, mediante el cual se transfiere información contenida en la secuencia de ADN hacia la secuencia de proteína utilizando diversos ARN como intermediarios.

<sup>3</sup> Monte Carlo es un método no determinístico o estadístico numérico, usado para aproximar expresiones matemáticas complejas y costosas de evaluar con exactitud.

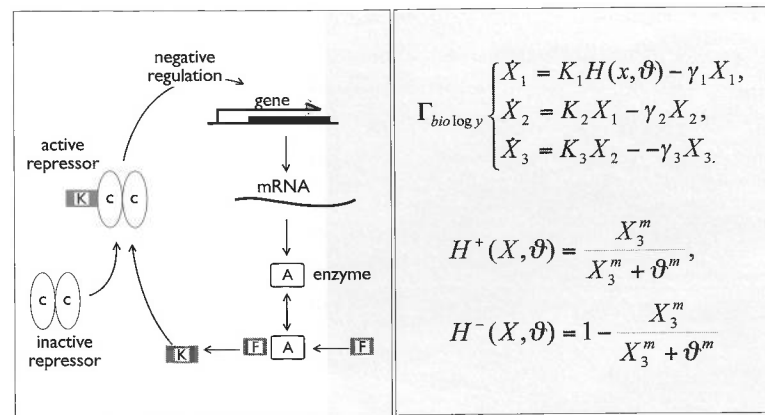


Figura 2 Modelo de un gen activado y retroalimentado por distintas moléculas. Para tres concentraciones  $X_1, X_2, X_3$  correspondientes al ARNm que codifica a una enzima inestable, la enzima y el metabolito respectivo, y donde  $K_1, K_2$  y  $K_3$  son las constantes de degradación, el modelo expresa el balance entre el número de moléculas que aparecen y desaparecen por unidad de tiempo en combinación de funciones no lineales. Fuente: Ibarra-Junquera et al. (2005) y Torres et al. (2007).

en un organismo, así como el momento preciso de su expresión. La regulación de la expresión es llevada a cabo por sistemas de regulación estructuradas por redes de interacción entre ADN, ARN, proteínas y pequeñas moléculas. Estas redes involucran muchos componentes conectados y retroalimentados, de tal manera que es de interés nuestro entender la dinámica entre sus señales. En consecuencia se utilizan métodos formales de simulación computacional, formalismos matemáticos y bioinformática. De esta manera sabremos como los genes son implicados en el control de procesos intracelulares y extracelulares. En la figura 2 se presenta un modelo sencillo de un gen interactuando con distintas moléculas. Para ello corresponde un modelo de ecuación que permite obtener la concentración de ARN (o RNA), y por consiguiente, conocer mediante un modelo la dinámica de expresión del gen. La ecuación representa el modelo obtenido para un solo gen (L. A. Torres et al., 2007).

De esta manera se modela un solo gen. Ahora, imaginemos que pasaría si queremos un arreglo de tres genes; entonces las ecuaciones se complican y por lo tanto el modelo se hace más complejo. Por otro lado, cuestionemos que pasaría en el caso de miles de genes, por ejemplo del cuerpo humano, donde se tienen ubicados alrededor de 30 000 genes y alrededor de 1 200 genes causantes de enfermedades. El problema se torna aun más complejo y es ahí donde la investigación desprendida de este proyecto genera modelos matemáticos y físicos (como el de Potts) para entender la interacciones de miles de genes que se encienden y apagan como focos de colores en un árbol de navidad; es así como se pretende escudriñar en el mundo nano y aportar al sector salud mediante modelos que indiquen a los médicos especialistas qué está ocurriendo en la expresión de genes al momento de padecer alguna enfermedad o conocer cómo un fármaco puede inhibir o activar determinado gen. Así mismo, se podrá obtener a través de la medicina genómica una medicina *individualizada*, donde el paciente podrá tener acceso a un medicamento hecho a su medida, es decir: ¡hecho a su genoma! Esa será la medicina en un futuro muy próximo...

### Conclusiones

La investigación interdisciplinaria será la clave en el siglo XXI para encontrar respuestas a múltiples enfermedades crónicas que actualmente inciden en nuestra sociedad y que aún no se tienen respuestas para encontrar una solución contundente.

La física, las matemáticas, la química y la biología son disciplinas fundamentales para abordar y fortalecer las nuevas ciencias como la bioinformática, la medicina genómica y la nanotecnología.

Los avances del proyecto de investigación han sido presentados en foros académicos internacionales con gran aceptación, además, de ser considerados como investigaciones interdisciplinarias a escala mundial y con enorme proyección en el futuro.

La intervención de estudiantes de ingeniería y nutrición en esta investigación es fundamental para la bioingeniería y en general para el desarrollo del país en todos los sectores industriales y tecnológicos.

Los algoritmos desprendidos de estas investigaciones pueden crear herramientas nuevas para detección del cáncer o en general de diversas enfermedades crónicas.

El análisis de los datos de microarreglos puede proporcionar información para el diseño de nuevos tratamientos, además de monitorear la eficacia del tratamiento mismo. ■

### REFERENCIAS

- Campbell, Malcom y Laurie J. Heyer (2003) *Discovering genomics, proteomics and bioinformatics*. San Francisco: Cold Spring Harbor Laboratory Press/Benjamin Cummings.
- Torres, Luis Adolfo (2008) "Genoma humano y bioética". En *Contexturas*, 28. Año VII. Diciembre-marzo. León: UIA León.
- V. Ibarra-Junquera, L.A. Torres, H.C. Rosu, G. Arguello, J. Collado-Vides (2005) "Nonlinear software sensor for monitoring genetic regulation process with noise and modeling errors". En *Physical Review E*, 72. 2 de octubre. Disponible en [http://arxiv.org/PS\\_cache/physics/pdf/0410/0410096v3.pdf](http://arxiv.org/PS_cache/physics/pdf/0410/0410096v3.pdf)
- V. Ibarra-Junquera, L.A. Torres, H.C. Rosu, G. Arguello, J. Collado-Vides. *Virtual Journal of Biological Physics Research* (2005) Vol.10, issue 3. Agosto.
- National Cancer Institute (2008) EUA. Disponible en [http://www.cancer.gov/Templates/db\\_alpha.aspx?CdrID=537335&lang=spanish](http://www.cancer.gov/Templates/db_alpha.aspx?CdrID=537335&lang=spanish)
- Schrödinger, Erwin (1944) *What is life? The physical aspect of the living cell*. Cambridge: University Press.
- Stekel, Dovel (2003) *Microarray bioinformatics*. Cambridge: University Press.