

Febrero de 2007

[http://www.biomedicas.unam.mx/noticias\\_gaceta.htm](http://www.biomedicas.unam.mx/noticias_gaceta.htm)

## Estructura y dinámica de redes genéticas

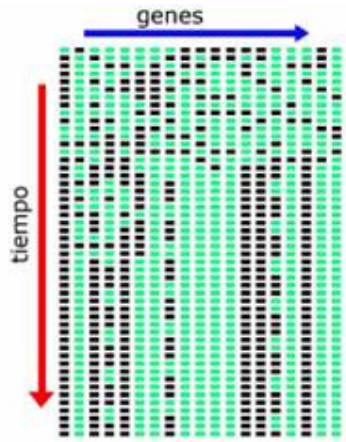
*Maximino Aldana González, Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México.*

*“Los organismos vivos son compatibles con las leyes físicas y químicas de interacción de la materia, pero no son una consecuencia de ellas”. Jacques Monod. *El azar y la necesidad* (1970).*

El nivel de complejidad y organización de la materia en los seres vivos es sorprendente, a la vez que fascinante. Aun cuando no sepamos por qué existe o cómo se genera, creemos firmemente en el paradigma que dice que gran parte de la información para el surgimiento de tal complejidad está codificada, de alguna forma, en los genes. Con el descubrimiento del código genético en la década de los sesenta, aprendimos que la información contenida en los genes determina la estructura de muchas de las proteínas (y ciertos tipos de ARN's no mensajeros) involucradas en el metabolismo celular. Sin embargo, pronto quedó claro que los genes sólo contienen la información para construir las piezas del rompecabezas, pero no contienen la información para armarlo, al menos no de manera directa. Así, incluso después de haber secuenciado el genoma completo de varios organismos, seguimos sin poder armar el rompecabezas de la vida, es decir, sin poder explicar cómo la información contenida en los genes determina la estructura, el funcionamiento y la gran complejidad de los seres vivos.

Aparentemente no hay suficiente espacio en las secuencias codificadoras de los genomas para albergar la información necesaria que determina las características fenotípicas de los organismos. Consideremos por ejemplo las funciones biológicas básicas que lleva a cabo un bebé recién nacido, tales como respirar, mantener su corazón latiendo, comer, llorar, abrir y cerrar los ojos, mover sus manos y sus pies, ingerir los alimentos y evacuar, responder a estímulos externos, etcétera. Estas funciones no son aprendidas, sino que forman parte del *programa genético* con el que nace todo bebé. ¿Cuántas neuronas y cuántas conexiones sinápticas entre ellas se necesitan para que el bebé lleve a cabo sus funciones biológicas básicas? No lo sabemos, pero tan sólo en el sentido del olfato, que hace que los bebés puedan reconocer el olor del pecho de su madre, están involucradas más de 40 millones de neuronas, cada una con 10 mil conexiones sinápticas en promedio. Pero lo que es más desconcertante es que no sabemos dónde está guardada la información de la arquitectura neuronal del cerebro de un bebé, es decir, dónde está el *blueprint* que determina la forma en que se deben establecer los millones y millones de conexiones neuronales para que el bebé sobreviva. Es claro que en los casi 26 mil genes del ser humano no hay espacio para almacenar de manera explícita esta información, más la información de los diferentes tipos celulares, más la

información de la secuencia temporal de las miles de reacciones químicas que ocurren dentro de la célula, etcétera. Y sin embargo, la evidencia experimental parece confirmar una y otra vez el paradigma genético: la información de gran parte de la estructura y funcionamiento de los seres vivos está contenida en los genes.



*Figura 1. Atractor dinámico. Cada cuadrado representa un gen, verde si está prendido y negro si está apagado, y cada renglón corresponde a un estado de expresión de todo el genoma. Comenzando con un estado de expresión inicial en el primer renglón, el genoma pasa por una serie de estados dinámicos hasta que se alcanza un patrón de expresión periódico denominado atractor.*

Pero el paradigma ha cambiado a través de los años. Durante décadas se estudiaron genes individuales, tratando de asociar un gen, o un grupo reducido de genes, con una característica fenotípica específica. Este enfoque reduccionista ha permitido entender ciertos fenómenos importantes, tales como la anemia falciforme en el ser humano, o las rutas lítica y lisogénica del bacteriófago lambda, fenómenos que pueden asociarse a un solo gen o a un *switch* genético compuesto de unos cuantos genes. Sin embargo, el estudio experimental y teórico de los genomas realizado en los últimos años ha evidenciado que la complejidad de los organismos vivos no es el resultado directo del número de genes, sino de los intrincados mecanismos de regulación de la expresión genética en todo el genoma. Bajo esta perspectiva, ciertas características fenotípicas, como el cáncer, sólo podrán comprenderse estudiando la estructura y la dinámica de *toda la red genética*.

Los genes interactúan unos con otros a través de las proteínas reguladoras (factores de transcripción) que generan. Los factores de transcripción se unen de forma específica a las regiones reguladoras de los genes, activando o reprimiendo su expresión. Esta serie de interacciones, mediadas por los factores de transcripción, da lugar a una red dirigida en la que dos genes están “conectados” si la expresión de uno regula, positiva o negativamente, la expresión del otro. El paradigma genético ha cambiado porque ahora sabemos que la información de las características fenotípicas del organismo, además de estar contenida explícitamente en la secuencia de bases de cada gen, también está contenida en la *topología* de la red genética (cómo están conectados los genes) y en su *dinámica* (qué genes se expresan en determinado momento). Estudios recientes muestran que entre más conectada está la red genética, más complejo es el organismo codificado en dicha red. Por ejemplo, el genoma de la mosca *Drosophila* codifica aproximadamente mil factores de transcripción distintos, mientras que el genoma de los seres humanos codifica por lo menos 3 mil factores de transcripción. El aumento en el número de factores de transcripción por genoma se traduce en una mayor conectividad de la red genética, lo cual a su vez está correlacionado con una mayor complejidad del organismo (M. Levine y R. Tjian. *Nature*, **424** (2003), p. 147). La red

genética, como un todo, puede albergar una cantidad de información mucho mayor que la que está explícitamente codificada en la secuencia de bases de los genes.

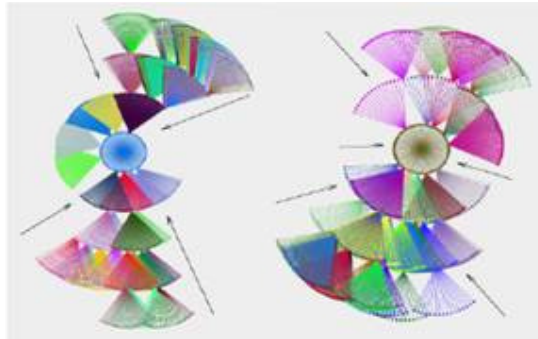
Calin Guet, de la Universidad de Chicago, ha llevado a cabo experimentos que demuestran que el funcionamiento celular puede cambiar drásticamente simplemente “recableando” algunas conexiones de la red genética. Calin demostró que es posible hacer que la bacteria *E. coli* sintética (o no) una proteína verde fluorescente simplemente intercambiando las regiones reguladoras de los genes *lacR*, *tetR* y *lambda cI*, sin cambiar las secuencias codificadoras en esos genes (C. Guet et al. *Science*, **296** (2002) p. 1466). Experimentos similares realizados por otros grupos de investigación han demostrado que la topología de la red contiene información importante de las características fenotípicas de los organismos. Por tal motivo, es importante determinar la topología de las redes genéticas de distintos organismos. Uno de los grupos que más ha contribuido a esta empresa es el de Julio Collado, director del Centro de Ciencias Genómicas de la UNAM, quien junto con sus colaboradores han construido la red genética más completa que se conoce hasta la fecha, la de la bacteria *E. coli* ([www.ccg.unam.mx](http://www.ccg.unam.mx)).

Por otro lado, la dinámica de la red genética tiene que ver con los estados de expresión de los genes a lo largo del tiempo. Sabemos que la expresión o no expresión de grandes conjuntos de genes determina características fenotípicas importantísimas tales como el tipo celular (hígado, pulmón, piel, etc.), el destino celular (apoptosis, duplicación, etc.), o en general el “estado funcional” de la célula. La dinámica genética que determina qué genes se expresan y qué genes no se expresan depende, fundamentalmente, de tres elementos: (a) la topología de la red y (b) la naturaleza activadora o represora de los factores de transcripción, y (c) las señales que recibe la célula de su contexto. Uno de los modelos más interesantes para estudiar la dinámica genética fue propuesto por Stuart Kauffman hace más de 35 años. En este modelo, cada gen se representa con una variable “ $g$ ” que sólo puede tomar dos valores:  $g = 1$  si el gen está expresado y  $g = 0$  si el gen no está expresado. Así, los  $N$  genes que conforman el genoma de un organismo se representa con una serie de variables discretas,  $g_1, g_2, g_3, \dots, g_N$ , que se “prenden” y “apagan” a lo largo del tiempo de acuerdo a ciertas reglas lógicas que se construyen con base en la naturaleza activadora o represora de los factores de transcripción de cada gen.

Tal vez el resultado más sobresaliente del modelo de Kauffman es la existencia de *atractores dinámicos*. Dada una configuración inicial del genoma, en la cual algunos genes se encuentran prendidos y otros apagados, se deja que la red evolucione dinámicamente en el tiempo. Cada gen, obedeciendo en todo momento a su regla lógica, pasa por una serie de estados prendido-apagado hasta que después de un tiempo no muy largo, toda la red entra en un estado de expresión cíclico. Algunos genes alcanzan un estado de expresión constante que ya no cambia en el tiempo, mientras que otros genes se siguen prendiendo y apagando pero de forma periódica (ver la figura 1). Este estado de expresión cíclico es el atractor dinámico de la red al que nos referimos antes. Generalmente, una misma red genética tiene varios atractores dinámicos. Cada atractor dinámico queda identificado de forma única por el conjunto de genes que están prendidos en él. En otras palabras, en atractores distintos se expresan conjuntos de genes diferentes. Y esta es precisamente la característica que identifica a los diferentes tipos celulares en un organismo: conjuntos distintos de genes se expresan en tipos celulares distintos. Por tal motivo, Kauffman formuló la hipótesis de que *los atractores dinámicos de la red genética corresponden a los diferentes tipos o destinos celulares*

*observados en el organismo.*

Esta hipótesis ha sido comprobada recientemente por varios grupos de investigación alrededor del mundo. Entre ellos destaca el grupo de Elena Álvarez Buylla del Instituto de Ecología de la UNAM. Basándose en datos experimentales, la doctora Álvarez Buylla y su grupo identificaron la estructura de la red de 15 genes que participan en el desarrollo floral de la planta *Arabidopsis thaliana*, así como las reglas lógicas que determinan el estado de expresión de estos 15 genes. Simulando la red en la computadora encontraron 10 atractores dinámicos, correspondientes a los estados de expresión observados experimentalmente en los distintos tipos celulares pertenecientes a los diferentes órganos de la flor (pétalos, estambres, carpelos, etcétera). Más aún, al cambiar en la computadora la estructura de la red “recableando” las conexiones de algunos genes se encontraron los patrones de expresión conocidos experimentalmente para los organismos mutantes (C. Espinosa Soto et al. *The Plant Cell*, **16** (2004) p. 2923). Reka. Albert y Hans G. Othmer obtuvieron resultados similares al aplicar el enfoque propuesto por Kauffman a la red de genes de polaridad de segmento de la mosca *Drosophila* (*Journal of Theoretical Biology*, **223** (2003) p. 1).



*Figura 2. Organización dinámica del espacio de expresión genética de una red artificial con 10 genes y  $2^{10} = 1024$  estados de expresión. Los estados de expresión están representados por puntos y se conectan con líneas si en la dinámica se siguen uno al otro en el tiempo (como los renglones consecutivos en la Figura 1). Las estructuras en forma de abanico reflejan que varios estados pueden converger a uno solo, y las flechas indican el sentido de la dinámica temporal. Para esta red en particular, los 1024 estados posibles se organizan en dos conjuntos, cada uno convergiendo a su propio atractor ubicado en el centro del círculo.*

La existencia de atractores dinámicos involucra la emergencia de un nivel de organización que no es en absoluto evidente a partir de las secuencias codificadoras en los genes. Para un genoma como el del ser humano con casi 26 mil genes, cada uno de los cuales puede estar prendido o apagado, existen cerca de  $2^{26000} = 10^{7826}$  estados de expresión del genoma completo, desde el estado donde todos los genes están apagados hasta el estado donde todos los genes están prendidos, pasando por todas las combinaciones intermedias de genes apagados y prendidos que se nos puedan ocurrir. El número de estados de expresión *posibles* del genoma humano es verdaderamente astronómico. Claramente, no todos estos estados son viables biológicamente. Pero lo fascinante es que los  $2^{26000}$  estados de expresión posibles se *organizan dinámicamente* en unos cuantos atractores, que se manifiestan en el fenotipo como los tipos o destinos celulares. Y no hay muchos atractores en una red genética. Simulaciones computacionales, realizadas en nuestro grupo de investigación, de la dinámica de la red de *E. coli*, que tiene aproximadamente

4 mil 300 genes, sugieren que en dicha red existen entre 100 y 200 atractores. Así, para esta red, el conjunto de los  $2^{4300} = 10^{1294}$  estados posibles de expresión del genoma se particiona en cien o doscientos conjuntos disjuntos, cada uno convergiendo a su propio atractor (ver la Figura 2).

La organización dinámica de los  $2^N$  estados de expresión de un genoma con  $N$  genes mostrada en la Figura 2 podría ser análoga a la organización neuronal en el cerebro del bebé. No tenemos que establecer explícitamente la información de cómo se conectarán todos y cada uno de los  $2^N$  estados de expresión para formar las estructuras mostradas en la Figura 2. Esta información es resultado de la dinámica temporal de la red genética, que con sólo  $N$  genes y con  $N$  reglas lógicas, una para cada gen, puede organizar y conectar dinámicamente un número astronómicamente mayor de estados. Es en este sentido que la dinámica de la red puede albergar una información mucho mayor que la explícitamente contenida en los  $N$  genes del organismo.

Estudiando teóricamente la dinámica de las redes genéticas hemos encontrado que la duplicación y posterior divergencia de un solo gen en todo el genoma puede generar nuevos atractores dinámicos, conservando con alta probabilidad los atractores que ya existían antes de la duplicación (M. Aldana et al. *Journal of Theoretical Biology*, (2007) doi:10.1016/j.jtbi.2006.10.027). Tomando en cuenta que los atractores representan características fenotípicas, este resultado significa la emergencia de nuevos fenotipos ante el proceso de duplicación y divergencia genéticas. Es decir, no sólo es el gen duplicado y mutado el que puede adquirir una nueva función, como se ha enfatizado en la literatura, sino que toda la red genética puede adquirir nuevas funciones y cambiar sus rutas de expresión ante la duplicación y divergencia de un solo gen. Las características fenotípicas codificadas en los nuevos atractores que emergen después de la duplicación pueden ser benéficas o no para que el organismo se adapte a nuevos entornos, pero en cualquier caso sirven como material sobre el que puede actuar la selección natural. Bajo este punto de vista, la evolución por saltos en la que fenotipos totalmente nuevos aparecen súbitamente, en oposición a la evolución gradualista en la que los fenotipos se modifican poco a poco, puede entenderse simplemente como la emergencia de nuevos atractores dinámicos ante una reconfiguración de la red genética.

En conclusión, estudiar las propiedades estructurales y dinámicas de las redes genéticas nos está permitiendo entender la enorme complejidad y diversidad de los organismos vivos. La red genética, como un todo, contiene información fenotípica que no es directamente discernible a partir de las secuencias codificadoras del genoma. Y esto es lo interesante del tema. Uno de los más grandes desafíos de la genómica moderna es diseñar técnicas experimentales, computacionales y analíticas que nos permitan extraer la información biológica contenida en la estructura y dinámica de toda la red genética. Sólo así podremos comenzar a armar el rompecabezas de la vida. ☘



ÍNDICE DE LA GACETA

2/2007