



La “materia oscura” de los sistemas biológicos

Juan Antonio García Ranea

Investigador Ramón y Cajal. Departamento de Biología Molecular y Bioquímica. Universidad de Málaga. ranea@uma.es

La materia oscura es un concepto surgido en la Astrofísica para definir aquella masa del universo que no podemos ver con los medios técnicos actuales, pero cuya existencia predecimos por la influencia indirecta que ejerce en el devenir de estrellas y galaxias.

En la era de la biología de sistemas hemos comenzado nuestra andadura en la investigación del cosmos de interacciones moleculares que articulan los sistemas vivos. En este contexto, y de forma análoga a lo ocurrido en la exploración del universo, los científicos comenzamos a intuir que grandes espacios del intrincado mundo de interrelaciones moleculares subyacentes a la vida aún permanecen oscuros a nuestro conocimiento.

La irrupción de la “revolución industrial” en los experimentos de biología molecular nos ha hecho creer que por fin contamos con los medios técnicos necesarios para la exploración de estos vastos espacios desconocidos. Pero la gran paradoja es que a pesar de que contamos con toneladas de datos experimentales no somos capaces de enunciar modelos sistémicos consistentes que expliquen la diversidad funcional de los sistemas biológicos complejos. Ante esta paradoja, cabría preguntarse ¿Cuál es la razón o razones que, en biología de sistemas, limitan tan significativamente la transformación de la información en conocimiento?

De forma general, la comunidad científica ha apuntado como principal sospechoso de dicha frustración a la falta de un mayor desarrollo formal y conceptual en una disciplina incipiente como es la Biología de Sistemas. No obstante, y sin menoscabo de que esta primera razón sea cierta, cabría también cuestionarse si con el conocimiento acumulado hasta la fecha, más las técnicas experimentales de exploración actuales, estamos en el buen camino para llegar a comprender como funcionan los sistemas moleculares que estudiamos. En resumen, podríamos decir que mientras la primera razón tendría que ver con una insuficiencia del desarrollo teórico de la biología de sistemas, la segunda apuntaría a un desarrollo ineficiente de la exploración experimental del espacio de interacciones, generándose así modelos incompletos difíciles de interpretar.

Adivinar qué progreso teórico futuro terminará por salvar las limitaciones actuales en este área de conocimiento se presenta como una tarea más especulativa que científica. En cambio, con los medios computacionales actuales sí es factible estimar el efecto que nuestras técnicas experimentales ejercen en el modelado de ciertos sistemas moleculares importantes, como por ejemplo lo son las redes de interacción entre proteínas. Esto es posible porque, al igual que en Astronomía la materia oscura se predice por su influencia indirecta en otros cuerpos celestes, muchas interacciones de las que aún no disponemos de evidencia experimental son producto de largos procesos de co-dependencia evolutiva entre proteínas que dejan su impronta en la estructura y función de los genomas, señales que son detectables mediante distintos métodos bio-computacionales.

Con el objeto de estudiar los males y virtudes de nuestras técnicas experimentales de exploración y estimar la suficiencia de los datos acumulados por la comunidad científica durante décadas, los grupos de la Profesora Christine Orengo, en la *University College London*, y nuestro grupo, en la Universidad de Málaga, colaboran en un estudio comparativo entre la red de interacciones de proteínas basada en el compendio de todas las evidencias experimentales disponibles en bases de datos públicas, y la red de interacciones predichas mediante métodos bio-computacionales [1]. Para dicho estudio utilizamos como objetos de investigación los proteomas completos de dos especies eucariotas muy diferentes, como son el hombre y la levadura *S. cerevisiae*. Especies que se consideran entre las mejor caracterizadas desde el punto de vista de la Biología Molecular entre los eucariotas, y a la vez lo suficientemente alejadas en el árbol de la vida como para poder ofrecernos la posibilidad de encontrar patrones universales en nuestros estudios. De esta forma

tenemos dos modelos independientes de la red de interacciones de proteínas que comparan en humano y levadura, por un lado una red compendio del conocimiento derivado de datos experimentales (red experimental), y por otro, la red de interacciones predicha mediante técnicas bio-computacionales (red de predicciones).

Con la comparación y estudio de ambas redes tratamos de dar respuestas a las siguientes preguntas: ¿Cuánto espacio de las redes de interacciones de proteínas permanece aún oscuro a nuestro conocimiento experimental? ¿Este espacio oscuro aún por explorar encierra componentes funcionalmente claves necesarios para comprender el funcionamiento del sistema? ¿Nuestras técnicas actuales de exploración experimental nos permitirán con el tiempo caracterizar todos los rincones del espacio de interacciones, o presentan limitaciones de muestreo que están dejando en la oscuridad importantes regiones funcionales?

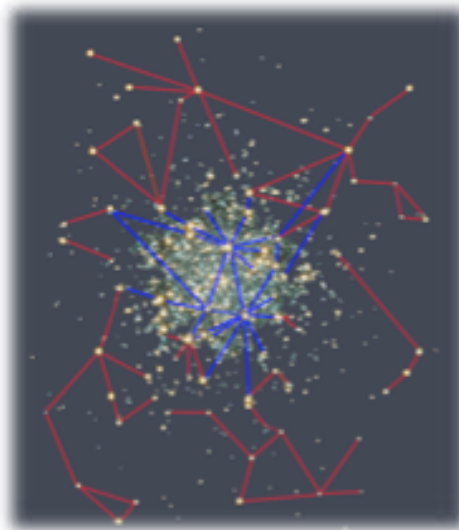
El estudio de solapamiento entre ambas redes [1] nos desveló que una proporción muy significativa de las redes de interacción entre proteínas predichas por medios bio-computacionales no ha sido aún caracterizada experimentalmente (el 82% en levadura y el 98% en humanos). A este espacio de interacción inexplorado lo denominamos la “materia oscura” de las redes de interacción. Aunque un alto porcentaje de materia oscura fue observado en ambos organismos, en la levadura el porcentaje de predicciones de las que sí tenemos evidencia experimental (18%) se estimó del orden de 14 veces superior que en humanos (1,34%), indicándonos que al menos desde el punto de vista de la suficiencia de los datos experimentales actuales, probablemente nos encontramos mucho más cerca de comprender el funcionamiento de los sistemas en *S. cerevisiae* de lo que estamos en el ser humano.

En la búsqueda de componentes funcionalmente claves en estos espacios no caracterizados experimentalmente, encontramos un alto número de proteínas *hubs* o nodos centrales de conexión, claves en la arquitectura de relaciones de la red. Al estudiar el papel funcional de estas proteínas que denominamos “*hubs* ocultos”, descubrimos que en la levadura correspondían principalmente a proteínas de membrana, mientras que en humanos estaban relacionados con funciones regulatorias en vías de señalización, como

las señales mediadas por proteínas quinasas mediante interacciones transitorias y que están implicadas en multitud de procesos de transmisión de señales y regulación celulares fundamentales de los organismos eucariotas.

Los resultados obtenidos en estos experimentos ponen en cuestión la manera en la que exploramos experimentalmente las redes de interacción molecular, así como nuestras posibilidades de llegar a modelar y comprender su funcionamiento. La estimación de la materia oscura nos indica que incluso para los organismos mejor estudiados como son el hombre y la levadura, en estos momentos estamos muy lejos de disponer de un catálogo suficiente de componentes y relaciones que nos permita generar modelos completos. Además, no solo nos falta mucha información, sino que desconocemos muchos componentes clave en el funcionamiento de dichas redes, como lo demuestra la existencia de nodos centrales o *hubs* ocultos. Por tanto, con datos incompletos e información determinante no disponible, es obvio concluir que nos encontramos muy lejos de poder generar modelos consistentes que nos permitan reproducir y entender el funcionamiento de dichos sistemas complejos [1,2].

La insuficiencia de datos no sería en sí demasiado preocupante si supiéramos que es sólo cuestión de tiempo que la aplicación recurrente de las técnicas experimentales actuales termine la caracterización de dichos espacios oscuros. Pero desafortunadamente, el análisis funcional de la materia oscura nos indica que nuestra exploración experimental de este espacio es sesgada, ignorando de forma sistemática regiones funcionalmente fundamentales, sin las cuales difícilmente alcanzaremos



a comprender el funcionamiento de los sistemas moleculares. Todos conocemos las limitaciones y el bajo rendimiento que distintas técnicas experimentales presentan en la caracterización, por ejemplo, de proteínas de membrana, las cuales están subrepresentadas en las bases de datos actuales, o las limitaciones que métodos como el sistema de doble híbrido en levadura presentan en la detección de interacciones transitorias como, por ejemplo, las que se producen en vías de señalización y cascadas de regulación protagonizadas por las proteínas quinasas tan abundantes en los proteomas de eucariotas complejos [3].

A la luz de los resultados actuales, podemos concluir que el avance del conocimiento en biología de sistemas no solo depende del desarrollo de nuevos conceptos, sino también de nuevos diseños experimentales de alto rendimiento que nos permitan corregir y compensar el sesgo de las técnicas aplicadas actualmente, así como completar la caracterización de las relaciones funcionales que permanecen oscuras y de las que por ahora sólo podemos tener evidencias indirectas.

La cosmovisión del hombre ha sufrido cambios dramáticos a medida que su conocimiento del universo se ha hecho más preciso, pasando de

un modelo geocéntrico a otro heliocéntrico, y posteriormente a la comprobación de la humilde ubicación de nuestro Sistema Solar en un brazo lateral de la galaxia Vía Láctea. Es probable que, de igual forma, se tenga que vencer la resistencia cultural a revisar alguno de los paradigmas y dogmas más queridos por parte de comunidades y grupos científicos bien establecidos en el ámbito de la Biología, lo que implicaría aceptar la idea de que probablemente las proteínas que se piensan centrales en muchos de los modelos experimentales actuales deban compartir su protagonismo con otras proteínas cuyo papel central no ha sido descubierto todavía. E incluso, en algunos casos, proteínas ahora consideradas claves deban dejar su papel protagonista a otras en modelos posteriores más precisos a medida que nos adentremos en la "materia oscura" de los organismos vivos.

Bibliografía citada:

1. Ranea JA, Morilla I, Lees JG, Reid AJ, Yeats C, Clegg AB, Sanchez-Jimenez F, Orengo C (2010). Finding the "dark matter" in human and yeast protein network prediction and modelling. *PLoS Comput Biol.* 6: pii: e1000945.
2. Wass MN, David A, Sternberg MJ (2011). Challenges for the prediction of macromolecular interactions. *Curr Opin Struct Biol.* 21: 382-90.
3. Russell RB, Aloy P (2008). Targeting and tinkering with interaction networks. *Nat Chem Biol* 4: 666–673.

65



Existo, luego soy una red

Albert Sorribas, Ester Vilaprinyó, Rui Alves

Departament de Ciències Mèdiques Bàsiques Institut de Recerca Biomèdica de Lleida (IRBLLEIDA) Universitat de Lleida

albert.sorribas@cmb.udl.cat



Obertura

Los seres vivos son altamente improbables desde el punto de vista de su complejidad. Si consideramos todos los procesos que acontecen en una simple célula, es difícil imaginar cómo ha llegado a ser posible. La respuesta es la evolución biológica mediante selección natural y otros procesos afines. Con ello, la explicación a este aparente misterio está servida, aunque todavía debemos emplearnos a fondo para comprender los mecanismos implicados y sus consecuencias. Desde este punto de vista, la Biología de Sistemas, que se caracteriza por una visión sistémica como marco conceptual para la comprensión de los procesos biológicos, su evolución y posible manipulación biotecnológica, surgió como una necesidad acuciante tras el desarrollo de tecnologías que abrieron un insospechado mar de nuevos datos y detalles a todos los niveles (genómico, estructural, metabólico, etc.).

Por lo tanto, en la biología de sistemas confluyen los siguientes problemas: (1) Una enorme capacidad experimental que permite caracterizar simultáneamente muchos componentes de distintos niveles de complejidad, (2) La necesidad de procesar esta información, (3) El reto de interpretar la complejidad, es decir, la necesidad de generar descripciones conceptuales que permitan entender dicha complejidad. La biología de sistemas conlleva un cambio de escala muy importante en la generación de datos experimentales y el reconocimiento de la importancia de las herramientas computacionales, desde la necesidad de buenas bases de datos y anotaciones, hasta la simulación de modelos, pasando por la reconstrucción de estructuras e interacciones moleculares.

En este corto ensayo nos centraremos en una idea central que subyace en la mayoría de cuestiones de la Biología de Sistemas: El concepto de red de interacciones (*network*) de muchos elementos.

El concepto de red de interacción

Uno de los avances importantes relacionados con la biología de sistemas es el desarrollo del análisis genómico a gran escala que ha permitido confirmar, como Darwin dejó claro, que el hombre no tiene un lugar especial en la evolución. Nuestra especie tiene un número de genes, entre 25000 y 30000, similar al ratón y a otras especies. En plantas, este número puede, fácilmente, triplicarse. Por otra parte, la mayoría de nuestros genes son idénticos en chimpancés.

Estos resultados de la genómica sacuden algunas ideas preconcebidas acerca del significado de la complejidad de las formas de vida. ¿Cómo puede ser que no tengamos muchos más genes que un ratón? La respuesta tiene que ver con la red de interacciones. No es lo mismo cuatro genes produciendo cada uno por su cuenta una proteína, que cuatro proteínas interaccionado dinámicamente en el espacio y el tiempo celular (desarrollo). La interacción de diversos elementos, como sucede en una sociedad humana, permite la emergencia de nuevas propiedades que no están presentes, ni son alcanzables, por los elementos individuales. Las técnicas experimentales han permitido caracterizar una parte de la complejidad de las interacciones a muchos niveles y permite entender que el número de genes no sea tan importante.

¿Cómo podemos entender las redes complejas?

Describir una red de interacciones es, hasta cierto punto, un trabajo rutinario porque disponemos de técnicas experimentales y computacionales que nos ayudan a procesar la información para obtener una abstracción útil. Así, por ejemplo, podemos representar las relaciones identificadas entre

todos los genes que, de una manera u otra, se relacionan con un cierto tipo de cáncer. O bien, podemos representar todas las relaciones que se dan entre los metabolitos y proteínas identificadas en un cierto tipo de células. ¿Y ahora qué?

La respuesta, como en muchos otros casos, depende en gran manera del tipo de problema que queramos resolver. Por ejemplo, podemos plantearnos cuestiones acerca de la similitud de la estructura de diversas redes a distintos niveles. ¿Podemos identificar propiedades básicas comunes? La respuesta es que sí y que redes tan distintas como el metabolismo, internet, o las relaciones sociales, parecen compartir un tipo de organización que se caracteriza por muchos nodos con pocas conexiones interaccionado con algunos nodos altamente conectados.

La situación es distinta cuando, una vez identificados los componente significativos de una red, nos planteamos resolver cuestiones como: ¿por qué una determinada persona desarrolla un tipo de enfermedad y otra, aparentemente similar, no? O, ¿por qué un tipo de tratamiento es efectivo en una persona y no en otra?

Las matemáticas al rescate

Aunque a algunos pueda sorprenderos, parte de la respuesta está en las matemáticas. El análisis intuitivo de situaciones complejas y la predicción de cambios dinámicos fruto de la interacción de muchos elementos no es, en general, posible sin el concurso de las matemáticas. El razonamiento intuitivo y lineal del tipo: Si A se activa, se activa B. Al interaccionar con B, C regula la expresión de..., tiene un límite porque las interacciones son no-lineales. Así, entender la progresión dinámica de los procesos celulares y predecir cómo se comportará la célula requiere el uso de herramientas matemáticas.

Los modelos matemáticos nos ayudan en muchas facetas prácticas, incluyendo el diseño de coches, el rendimiento deportivo, el análisis económico, etc. La biología de sistemas ha situado, entre otros avances, a los modelos matemáticos en el centro de las herramientas necesarias para entender la biología.

Modelos: una pasarela para entender la complejidad

Cuando el problema es complejo, necesitamos captar la esencia del mismo y representarlo en un modelo conceptual (matemático) útil que permita su solución. En este punto, la biología presenta muchos problemas metodológicos. Por una parte, el número de elementos de la forma de vida más simple es enorme. Necesitamos representar sus interacciones (que en gran parte son desconocidas) de una forma que podamos predecir qué pasará en respuesta a ciertos estímulos. Además deberíamos ser capaces de identificar qué principios de organización son comunes a diferentes redes, fruto de la evolución.

Actualmente, los modelos matemáticos son fundamentales en Biología de Sistemas. Por ejemplo, los modelos de ciclo celular han permitido comprender mejor el control de las distintas fases del ciclo. La importancia de este conocimiento para entender los procesos cancerosos es evidente y permite apreciar la importante contribución de esta técnica en la realidad de la investigación aplicada. A otro nivel, se están realizando grandes esfuerzos para modelar órganos humanos, por ejemplo el hígado, como un paso fundamental para disponer de un laboratorio barato donde probar nuevas terapias. En este sentido, el uso de modelos matemáticos puede ser un factor muy importante que contribuya a reducir el coste de desarrollar nuevos tratamientos, beneficiando a los pacientes con un menor coste de comercialización.

El cuello de botella en toda esta actividad es la obtención de datos fiables acerca de los componentes celulares que permitan una modelización realista. Las cuestiones técnicas relacionadas con la simulación por ordenador, la aplicación de técnicas de optimización de procesos para evaluar el efecto de cambios en el sistema, etc., pueden darse por resueltas. La capacidad actual de cálculo de los ordenadores permite afrontar grandes proyectos sin demasiados problemas. Es de esperar que en un futuro próximo, las técnicas experimentales den un salto cualitativo y permitan la obtención de dichos datos con los que plantear la solución de problemas de gran interés médico.

¿Qué promesas podemos cumplir en un futuro próximo?

En los primeros tiempos de la aceptación general del concepto de Biología de Sistemas, hace unos diez años, se produjo un cierto estallido de predicciones. Entre ellas, la posibilidad de caracterizar el genoma de cada persona abre grandes posibilidades médicas al permitir identificar genes alterados. Es evidente que la ausencia o disfunción de un gen determinado puede facilitar el desarrollo de una patología, como es el caso del gen BRCA1 en el cáncer de mama. Sin embargo, en muchos casos dichas alteraciones provocan disfunciones que pueden compensarse, en parte, por el incremento de función de otros genes o procesos. Esto dificulta relacionar la patología (fenotipo) con la genética de cada individuo.

Aunque los esfuerzos de investigación son muy importantes, la comprensión del concepto de red nos hace ser prudentes hacia la posibilidad de desarrollar, en un futuro inmediato, una medicina verdaderamente personalizada, por falta de información acerca de los procesos celulares implicados en una enfermedad. Sin embargo, disponemos de algo valioso. Sabemos lo que nos falta. Por lo tanto, se trata de poner manos a la obra y de ser osados. Cuando se propuso el proyecto genoma humano muchos dijeron que era una tarea excesivamente compleja. Al final, el proyecto se completó en un tiempo récord, dando paso al interés general por la Biología de sSistemas.

De la descripción de la vida a la creación de nuevos organismos

Uno de los aspectos prácticos importantes de la biología de sistemas es la posibilidad de modificar organismos (biotecnología) o de crear nuevos organismos y introducir circuitos metabólicos diseñados para realizar determinadas funciones (Biología Sintética). En ambos casos, nos encontramos ante nuevas posibilidades que, como muchos casos en ciencia, abren grandes interrogantes éticos aunque pueden proporcionar grandes avances prácticos. La biotecnología basada en un análisis racional de las propiedades de los procesos celulares (ingeniería metabólica) debe permitir la obtención de organismos modificados que incrementen, por ejemplo, la producción de determinados productos de interés práctico. El ahorro en experimentos, animales de laboratorio, etc., puede ser muy importante. En este sentido, la biotecnología es ya una de las actividades económicas más dinámicas y con un futuro muy interesante.

Mensaje

La biología de sistemas tiene ya resultados espectaculares en diversos ámbitos, pero plantea retos tan amplios que se necesitarán años para lograrlos. Si os seduce la idea de trabajar en este ámbito, interesaros por la biología, la computación, las matemáticas, ... ¡y sed muy curiosos!

Lecturas recomendadas para saber más:

Guimerà R, Sales-Pardo M (2009) Missing and spurious interactions and the reconstruction of complex networks. *Proceedings of the National Academy of Sciences USA* 106(52):22073-22078.

Koonin EV (2012) *The logic of chance: The nature and origin of biological evolution*. FT Press Science. New Jersey, USA.

Nalylor S, Chen JY (2010) Unraveling human complexity and disease with systems biology and personalized medicine. *Personalized Medicine* 7(3):275-289.

Scott Turner J (2007) *The tinkerer's accomplice: How design emerges from life itself*. Harvard University Press. Cambridge, Massachusetts, USA.

