

**EDITORIAL
DE LA BIOLOGÍA MOLECULAR
A LA GENÓMICA**

EDITORIAL

DE LA BIOLOGÍA MOLECULAR A LA GENÓMICA

La biología molecular tiene sus antecedentes en dos escuelas: la escuela estructural y la escuela informacional. Ambas tienen sus orígenes en las décadas de los años 30's, 40's y 50's del siglo pasado. La escuela estructural se preguntaba si sería posible inferir la función de las macromoléculas a través del conocimiento de su estructura tridimensional. Este es un enfoque muy utilizado en el desarrollo de la biología y ha servido múltiples veces. Por ejemplo, los grandes anatomistas lograron entender parte del funcionamiento del organismo a través del conocimiento de las estructuras anatómicas; los microscopistas renovaron la visión de la naturaleza a través de la visualización de microorganismos nunca antes vistos; el mismo sistema taxonómico propuesto por Linneo se basa en la observación de las morfologías de los organismos; y por supuesto el trabajo monumental de Darwin se desarrolló a partir de sus observaciones, en particular las hechas en las islas de las Galápagos sobre la variación en la forma del pico de los pinzones. Dicho de otro modo, mucho del avance de la biología está sustentado en observación—cuantificación—medida—inferencia.

En el caso de la biología molecular, la estrategia para conocer las estructuras moleculares fue la difracción de rayos X desarrollada inicialmente por William Henry y William Lawrence Bragg,

padre e hijo respectivamente, y que compartieron el premio Nobel de Física por el desarrollo de estas tecnologías. Quienes usaron inicialmente la cristalografía de rayos X en proteínas fueron Max Perutz y John Kendrew y resolvieron las primeras estructuras tridimensionales de proteínas; en concreto, de la hemoglobina y la mioglobina.

Resultó que la posibilidad de inferir directamente la función a través de la visualización de la estructura fue algo más complicado, ya que, como mencionaban en aquella época, las cadenas polipeptídicas parecían salchichas retorcidas y no era evidente a primera vista cómo comprender su función. Esto tomó más tiempo pero, efectivamente, durante los años posteriores el análisis de muchas estructuras tridimensionales de proteínas ha ido revelando los detalles de su función a partir de sus estructuras tridimensionales. Como sea, la escuela estructural planteaba que la física tenía mucho que aportar a la biología a través de la difracción y la cristalografía de rayos X.

Por otro lado, la escuela informacional se preguntaba cuál sería la naturaleza física del gen; cómo explicar a nivel molecular la transmisión de los caracteres hereditarios de padres a hijos; cómo se manifiesta y se regula, a nivel molecular, el flujo de información genética en cada individuo. Dicho de otro modo, se estaban preguntando cuál era la relación, a nivel molecular, entre

genotipo y fenotipo, filogenia y ontogenia, o para decirlo en términos modernos, cómo se relacionan el genoma, el transcriptoma, y el proteoma.

El origen de esta escuela es curioso y se remonta al momento en que el físico atómico Niels Bohr, desviándose un poco de su campo tradicional de investigación, en los años 30's del siglo XX, empezó a reflexionar acerca de la biología y en particular, sobre la naturaleza más esencial del fenómeno hereditario. Estas disquisiciones llevaron a Bohr a proponer un principio de incertidumbre biológica *grosso modo* equivalente al principio de incertidumbre cuántica de la física. El razonamiento de Bohr tiene que ver con los trabajos realizados a finales del siglo XIX por el bioquímico Eduard Buchner en los que demostró que en un extracto de levaduras libre de células era posible que se llevaran a cabo reacciones químicas de fermentación; esto es, reacciones enzimáticas en ausencia de células. Éste es uno de los grandes logros intelectuales de la ciencia ya que demuestra que los sistemas biológicos funcionan bajo las mismas leyes físicas y químicas que el resto de la materia. Sin embargo, Niels Bohr vio estos resultados de una manera diferente y razonó que para estudiar algo, es necesario respetar las características fundamentales del objeto en cuestión. En el caso particular de la vida, hay que respetar la característica más esencial del fenómeno vital que es, precisamente, estar vivo; y por lo tanto decía que lo que estudiaba Buchner podría ser cualquier cosa, pero no vida porque, efectivamente, en el tubo de ensayo ya no había vida.

Dicho de otro modo, Bohr pensaba que para estudiar un sistema, a éste se le debe de permitir conservar las características que lo definen, en este caso estar vivo. Como conclusión, propuso que debieran de existir otras leyes de la física en la biología, necesarias para comprender que el fenómeno vital y, en particular, el proceso hereditario no podrían ser explicados con las mismas leyes físicas y químicas que se aplican al resto de la materia. *Obviamente, estaba cayendo en el terreno del vitalismo.* Estas ideas las plasmó en su escrito "*Life and Light*". Pero por supuesto es más fácil proponer algo que demostrarlo y no fue Bohr quien se lanzó a la búsqueda de las nuevas leyes de la física en la biología sino un discípulo suyo, Max Delbrück acompañado inicialmente por Salvador Luria. El enfoque de Delbrück y Luria fue buscar el sistema biológico más simple que tuviera las características que querían estudiar. Se daban cuenta de que los organismos modelos

que se habían usado anteriormente eran sumamente complejos y decidieron usar bacteriófagos, los cuáles no están vivos, pero presentan el fenómeno de transmisión de sus caracteres hereditarios. Así se conocieron como el "Grupo del Fago" que, entre otros de sus integrantes tuvo a James Watson (discípulo de Luria). Como era de esperarse, no descubrieron ningunas otras leyes de la física en la biología, pero sus estudios sobre la infección de las bacterias por los fagos establecieron las bases de lo que sería la genética microbiana de la segunda mitad del siglo XX y ambos recibieron el premio Nobel de Fisiología o Medicina por sus aportes. Entre estos se podría enfatizar que su prueba de fluctuación demostró que la selección natural operaba también en bacterias. Curiosamente, uno de los mayores errores conceptuales del Grupo del Fago fue pensar que las proteínas y no los ácidos nucleicos son el material de la herencia. Y es interesante que hayan cometido este error porque no desconocían los trabajos de Oswald Avery y sus colaboradores quienes, en 1944, demostraron que los genes en las bacterias están constituidos de ácido desoxirribonucleico (DNA). En 1952, Alfred Hershey y Martha Chase, miembros del Grupo del Fago, demostraron que lo mismo ocurría con los bacteriófagos.

Otro físico que se asomó a la biología en aquellos años fue Erwin Schrödinger, quien, aunque no sabía de los ácidos nucleicos, en su libro de 1944 "*What is Life?*" propuso que el gen debiera ser un cristal aperiódico independientemente de su naturaleza física. Si se piensa un momento, el modelo de la doble hélice propuesta por James Watson y Francis Crick en 1953, efectivamente podría considerarse como un cristal aperiódico. En 1971, Jacques Monod, quien conocía el libro de Schrödinger, escribió "el gen es un cristal aperiódico en el sentido de que la secuencia no es repetitiva y totalmente libre. La estructura -de la doble hélice- puede acomodar todas las secuencias posibles".

Pero para llegar al modelo de la doble hélice faltaría mencionar un antecedente importante, tan importante como los trabajos de difracción de rayos X, y que es el análisis de la composición cuantitativa de las bases nitrogenadas que llevó a cabo Erwin Chargaff. No se puede disminuir la relevancia que tuvieron sus resultados al demostrar que la cantidad de adenina es igual a la de timina y que la cantidad de guanina es igual a la de citosina. Estos datos llevaron a Watson y Crick a proponer que debiera de haber un apareamiento

to de bases A-T/G-C y que este apareamiento podría explicar cómo es que la información genética se copia y pasa de generación en generación. Sin los resultados de Chargaff, Watson y Crick hubieran podido proponer un modelo similar, pero hubiera quedado indefinido en el sentido de que cualquier purina se podría aparearse con cualquier pirimidina y el impacto del modelo hubiera quedado muy restringido. Además, Watson y Crick se dieron cuenta de que la secuencia no tenía restricciones siempre y cuando se mantuvieran los apareamientos complementarios A-T/G-C y que por lo tanto en la secuencia de nucleótidos podría haber un código para cifrar la secuencia de aminoácidos de las proteínas. Dicho de otro modo, el modelo explica la transmisión de los caracteres hereditarios de padres a hijos (filogenia) y cómo es que esa información determina las características fenotípicas de un sistema biológico (ontogenia), con lo cual se concretaron las preguntas de la escuela informacional.

A partir de la propuesta del modelo de la doble hélice, los avances fueron vertiginosos. En menos de 10 años, Paul Zamecnik y Mahlon Hoagland establecieron que los ribosomas son el sitio donde se lleva a cabo la síntesis de proteínas y describieron al RNA de transferencia; Arthur Kornberg describió a las DNA polimerasas y resolvió los fundamentos de la replicación; se fueron resolviendo las primeras estructuras tridimensionales de proteínas y Frederick Sanger obtuvo la primera secuencia de residuos de aminoácidos de un polipéptido y con esto se abrió el capítulo de la evolución molecular. Posiblemente una de las mayores extensiones del modelo de la doble hélice fue la propuesta de François Jacob y Jacques Monod del modelo del operón que definió el paradigma de la regulación de la expresión génica y además, de manera colateral, llevó a elucidar el código genético. En resumen, en unos cuantos años a partir del modelo de Watson y Crick se entendió de manera general el funcionamiento de los sistemas biológicos a nivel molecular.

Hace 50 años la biología molecular había llegado a su fase académica y en términos generales se entendía cómo se daban los flujos de información entre las macromoléculas informacionales, la esencia de la regulación génica, y se percibía ya cuáles eran las fuentes de la evolución molecular. El impacto de este conocimiento en lo que habría de venir fue enorme. Ciertamente había algo de arrogancia -los secretos básicos habían sido develados- y sólo faltaban algunos detalles por conocer. Pero la biología molecular todavía tenía

mucho por delante y a mediados de los años 70's del siglo XX, se describieron tecnologías sumamente poderosas que en general conocemos como clonación molecular, ingeniería genética o DNA recombinante, las cuales catapultaron las posibilidades de análisis a nivel molecular. Así, poco a poco aparecieron los genes fragmentados, los códigos genéticos mitocondriales y los organismos genéticamente modificados. Entre estas tecnologías hay una que vale la pena destacar: la posibilidad de conocer la secuencia de nucleótidos en los ácidos nucleicos y, por extensión, por el código genético conocer la secuencia de residuos de aminoácidos en las proteínas.

Las técnicas de secuenciación de DNA desarrolladas por Fred Sanger y sus colaboradores ejemplifican cómo la tecnología y el cuerpo formal de conocimientos se retroalimentan mutuamente. Desde que Watson y Crick propusieron el modelo de la doble hélice, quedaba claro que era interesante conocer la secuencia de los nucleótidos. Pero en los primeros años, las técnicas desarrolladas eran sumamente limitadas y burdas. Por eso, al inicio, Sanger le dio la vuelta al problema y decidió hacer la secuenciación de proteínas que, por así decirlo, de alguna manera es secuenciar a los genes de manera indirecta (de segunda mano). La química del momento permitía estas metodologías para proteínas, pero no para los ácidos nucleicos. Sanger fue paciente y a finales de los años 70's del siglo XX publicó las estrategias de secuenciación que duraron décadas.

Lo que vale la pena enfatizar es el razonamiento de Sanger. Apoyándose en lo que se había avanzado en el conocimiento de la síntesis de DNA, a Sanger se le ocurrió que quien "sabe" la secuencia de nucleótidos es la DNA polimerasa (DNA Pol), en el sentido de que está siguiendo las instrucciones del templado. Por lo tanto, se trataba de diseñar una metodología que nos dijera el orden de cada uno de los nucleótidos incorporados por la DNA polimerasa. En eso radicó el genio de Sanger y lo que le dio su segundo premio Nobel de Química, en 1980. Inicialmente, la estrategia que desarrolló consistía en un ingenioso ensayo de presencia/ausencia de cada nucleótido que permitía deducir el orden de incorporación de los nucleótidos por la DNA Pol. Lamentablemente, el ensayo era complicado y no libre de errores, por lo que después desarrolló el método de los didesoxinucleótidos que provoca la terminación de la polimerización cuando el nucleótido que se incorpora es uno de los cuatro

didesoxinucleótidos. Actualmente, existen muchos métodos de secuenciación eficientes, confiables, y rápidos, pero la genómica se inició con los métodos de Sanger y no hay que olvidar que los primeros borradores del genoma humano se hicieron con las estrategias de Sanger.

El primer genoma totalmente secuenciado por Sanger y sus colaboradores, fue el del bacteriófago Phi-X-174. Su tamaño era de sólo 5,375 pares de bases, pero la emoción de haber logrado este conocimiento era varios órdenes de magnitud mayor. A continuación, poco a poco fue aumentando el tamaño de los genomas secuenciados, pasando del genoma mitocondrial, el fago Lambda, fagos de tipo T4, hasta que finalmente el grupo de Craig Venter, en 1995, reportó el primer genoma bacteriano (*Haemophilus influenza* ~ 2 Mega pb). El resto es historia hasta llegar al genoma humano (3 Giga pb), a finales del siglo XX.

Durante el presente siglo, los proyectos de secuenciación de genomas se van volviendo más y más ambiciosos cuantitativamente –no necesariamente cualitativamente. Por ejemplo, el “*Vertebrate Genomes Project*” pretende secuenciar a las aproximadamente 66,000 especies de vertebrados existentes en un lapso de algunos pocos años. Ni qué decir de los proyectos de secuenciación de DNA antiguo. Pero más cantidad de información en las bases de datos no necesariamente significa que tengamos una mejor comprensión de su significado. Simplemente diré que aún hoy en día no entendemos totalmente el funcionamiento del primer genoma secuenciado (Phi-X-174, que tiene entre 7 y 9 genes).

La genómica ha impactado muchas áreas, por ejemplo, la medicina genómica, la agrogenómica, la farmacogenómica, etc. Pero uno de los mayores impactos de la secuenciación de genomas está en el terreno evolutivo. En particular, la historia evolutiva del hombre develada por el registro fósil ha permitido reconstruir el devenir de los homíninos durante los últimos 6 millones de años, esto es a partir del último ancestro común con el chimpancé. De igual manera, haciendo comparaciones de la secuencia de nucleótidos de los genes o de aminoácidos de las proteínas podemos hacer comparaciones de nuestro genoma con el de muchísimas especies y ubicarnos en esa constelación genómica como una pequeña estrella. No somos la especie más importante ni más relevante. Somos, igual que el resto de los organismos contemporáneos, el resultado de un proceso continuo de evolución de más de 3,500,000,000 de años.

Uno de los resultados más concretos de la secuenciación del genoma de diferentes individuos de nuestra especie es que, entre dos humanos cualquiera, independientemente de su origen geográfico tenemos un 99.9% de identidad en la secuencia. Este dato derrumba el concepto de raza. Esta diferencia molecular parece muy pequeña, pero en realidad es mucho mayor, porque si se comparan más individuos las secuencias específicas que determinan el 0.1% de diferencia no necesariamente son las mismas. De hecho, análisis de las secuencias codificantes de más de 60 mil humanos revelan que, en conjunto, hay una diferencia cada ocho pares de bases. Dicho de otro modo, la variabilidad del genoma humano a nivel poblacional es enorme. Justamente es lo que Darwin postuló: la variabilidad (que se hereda), más la selección natural, conducen el proceso evolutivo de especiación.

Pero también sabemos que con los chimpancés actuales tenemos un 99% de identidad genómica. Para explicar este alto parecido molecular y reconciliarlo con las diferencias anatómicas y morfológicas obvias (~ 30%), Allan Wilson y Marie Claire King propusieron que, independientemente de que los dos genomas son muy similares y tenemos prácticamente los mismos genes, el nivel de expresión no es el mismo y es esto lo que determina las diferencias anatómicas, morfológicas, bioquímicas o fisiológicas. Una plétora de análisis con una variedad de genes en ambas especies, confirman hoy la propuesta de Wilson y King. Un ejemplo sencillo puede ayudar: en ambos genomas se localizan los genes que determinan la síntesis del cabello; sin embargo, de manera obvia, en el caso del humano el nivel de expresión está muy disminuido. La explicación de esta característica es que los humanos tuvieron que adaptarse a un medio ambiente muy diferente al de los chimpancés y además adquirieron otra característica que es la de ser corredores de distancia en África y por lo tanto la termorregulación requirió de cambios en varios niveles, uno de los cuales es la disminución del vello corporal.


Hoy también sabemos de varios genes que sufrieron cambios adaptativos en los últimos millones de años. Por ejemplo, hace tres millones de años aproximadamente, un gen que regula el ciclo de diferenciación y proliferación en el desarrollo del cerebro tuvo cambios que llevaron a un aumento en el número de neuronas, mientras que cambios en otro gen que regula la maduración de las dendritas llevó a que se

establecieran más sinapsis. Más adelante, hace aproximadamente dos y medio millones de años, uno de los genes de miosina muscular sufrió una inactivación lo cual provocó una pérdida de la fuerza de la masticación. Es interesante que, en vez de generar una desventaja, este cambio provocó que nuestros ancestros cambiaran de una dieta predominantemente vegetariana, a una dieta omnívora, lo cual condujo a la obtención de una mayor cantidad de energía y nutrientes. No puede dejar de mencionarse que la sofisticación en el uso del lenguaje articulado depende de otro gen muy especial (el FOXP2) asociado a éste.

La comunicación es una característica antigua en los vertebrados, pero a partir del linaje humano han ocurrido cambios notables en un gen asociado al lenguaje, dos de los cuales ocurrieron hace aproximadamente 600,000 años y que coinciden con las etapas finales de transformación del lenguaje articulado. Durante un tiempo y basándose en los restos del aparato fonador, se debatió si los neandertales usaban el lenguaje articulado. Hoy, con la secuencia del genoma de los neandertales, sabemos que portaban la misma versión del gen asociado al lenguaje que nosotros. Por cierto, hoy también

sabemos que los humanos contemporáneos y los neandertales tuvieron un encuentro reproductivo que llevó a que en el genoma de la mayor parte de los humanos existan huellas moleculares de aquellos intercambios genéticos. Solo para mencionarlo, el porcentaje de identidad genómica entre estos dos grupos de humanos es de 99.7%.

Hoy, por lo demás, la evolución continúa y claramente ha actuado en varios niveles en las poblaciones humanas. La enorme diversidad que muestra nuestra especie es consistente con la diversificación y la adaptación. De hecho, sabemos que hay varios genes que han estado sujetos a presión de selección; por ejemplo, entre los habitantes del Tíbet, en los nómadas del mar de Indonesia o en los pueblos amazónicos, por mencionar algunos.

Para concluir, diré que, en el devenir de la biología molecular a la genómica, en los últimos 80 años aproximadamente nos hemos acercado a comprender el funcionamiento de los sistemas biológicos a nivel molecular y nos hemos aproximado a vislumbrar algo de sus procesos evolutivos, pero todavía falta un largo camino hacia adelante. 

Dr. Víctor Valdés López
Laboratorio de Biología Molecular y Genómica
Departamento de Biología Celular.
Facultad de Ciencias, UNAM.
Editor de la REB
vvaldes@unam.mx

Adrián Romero-Chaveste