

Lección inaugural del curso 2000-2001

Genomas y proteomas

JOSEP M. SERRAT I JURADO

Una lección como la que hoy tengo el placer de leer con motivo del acto inaugural del curso 2000-2001 tiene que tener forzosamente unas características algo diferentes de las lecciones que impartimos a lo largo del curso en las aulas de nuestra universidad. Ustedes provienen de campos del conocimiento muy diversos y no quisiera cansarles, sobretodo a la audiencia menos familiarizada con los conceptos de las ciencias naturales, con un tema de especialización bioquímica. La lección de hoy quiere tener un carácter general, aunque esto comporte tener que dejar algunos temas abiertos y que otros no se puedan profundizar como se haría en una clase convencional. A la hora de preparar esta lección, la intención por mi parte ha sido explicar de un modo sugerente y lo bastante amplio el mundo de la información genética, cuyo protagonista es el ADN, el ácido desoxiribonucleico.

Muchas de las propiedades que hacen del ADN una molécula tan especial se pueden explicar sin que tengamos que hacer previamente una descripción a fondo de su estructura y comportamiento químico; de hecho, algunas de las características funcionales del ADN se han descubierto sin la intervención de las herramientas experimentales de la bioquímica¹. Sabemos que el ADN se puede representar mediante una secuencia lineal de cuatro letras, la C, la G, la T y la A, que en el vocabulario especializado se llaman abreviadamente bases o nucleótidos. Cada una de estas letras simboliza un compuesto químico diferente. El ADN está formado, pues, por una cadena de moléculas; en términos químicos se trata de un polímero. La información genética de un determinado individuo es el «texto» que se podría leer en su ADN completo, su genoma. En el caso de los seres humanos este texto tiene aproximadamente tres mil millones de caracteres. El famoso Proyecto Genoma Humano, del que hablaremos más adelante, tiene por objetivo saber qué dice el ADN de los humanos desde la primera hasta la tres mil millonésima letra

Pero, ¿qué debemos entender por «información genética»? ¿De qué informa el genoma? Para explicarlo permítanme hablar antes de otro polímero biológico de una importancia capital para el funcionamiento de los seres vivos: las proteínas.

Todos estamos familiarizados con el término «proteína» y con frecuencia lo asociamos

1. Robert Holliday lo expresaba de la siguiente manera: “los geneticistas [de los años 60] tendían a pensar en términos conceptuales más que en términos bioquímicos o moleculares. El ADN, para ellos, era una estructura física que podía mutar y recombinarse y no había que preocuparse, en aquel estadio, de la bioquímica subyacente.” (Holliday, 1990).

con los alimentos que consumimos diariamente. No obstante, la función de estas moléculas en el organismo sobrepasa en mucho la mera función nutritiva. De hecho, la mayor parte de seres vivos no las consumen nunca. Para percatarnos de la importancia de las proteínas quizá nos ayudará presentar aquí una pequeña muestra: las estructuras macroscópicas de los animales, como el esqueleto, la musculatura, las plumas de las aves, la piel, el pelo y las uñas son proteínas; las enzimas, que catalizan todas las reacciones bioquímicas, también son proteínas; los anticuerpos, la hemoglobina de la sangre, la insulina, la clorofila de las plantas, la cápside de los virus, los cilios y flagelos de los microorganismos, los antígenos responsables del rechazo de órganos en los trasplantes, las moléculas que regulan el transporte de sustancias a través de la membrana celular y los agentes patológicos que causan la enfermedad de las vacas locas, los priones, son también proteínas. A lo largo de la vida de un ser humano actúan aproximadamente cien mil proteínas diferentes². No sería atrevido decir que las proteínas hacen que seamos lo que somos.

¿Cuál es pues la relación entre el ADN y las proteínas? Las proteínas son también un polímero lineal, en este caso formado por veinte tipos diferentes de moléculas, los llamados aminoácidos, que también se pueden representar mediante una letra³. Pues bien, la secuencia de aminoácidos de una proteína se puede traducir mediante un código, el «código genético», en una secuencia del alfabeto de cuatro letras del ADN. La secuencia resultante forma parte de la información genética del individuo y se encuentra inserta en algún lugar de su genoma. Este segmento se denomina el «gen» de la proteína⁴. Pero lo que pasa en realidad es el proceso inverso. La célula dispone de un mecanismo químico para poder traducir la información del ADN y sintetizar todas las proteínas que necesita para su metabolismo.

Por tanto, cuando hablamos de información genética queremos decir información para poder sintetizar proteínas. El ADN aporta la información y las proteínas realizan las funciones necesarias para mantener con vida a un ser vivo. Ahora bien, como veremos enseguida, el mensaje genómico contiene además otros tipos de información.

Los seres vivos estamos formados por células. Cada célula dispone de una cantidad fija de material genético, la misma cantidad y la misma secuencia de bases en todas las células de un mismo individuo, tanto si se trata de las células que forman el tejido adiposo como si se trata de las neuronas de la corteza cerebral. Lo que diferencia a los tejidos a nivel molecular es el conjunto de

2. Esta cantidad es la estimación que se hacía antes de la secuenciación completa de los cromosomas humanos 21 y 22. El análisis del ADN de estos cromosomas indica que el número de genes de un ser humano podría ser menos de lo que se había estimado (Dunham *et al.*, 1999; Hattori *et al.*, 2000). No obstante, aún no se puede decir nada. Los cromosomas 21 y 22 son los más pequeños de los humanos, sin contar los cromosomas sexuales. Para poder conocer con más exactitud este dato tendremos que esperar a la secuenciación del resto de cromosomas. Hay que tener en cuenta, además, que la relación entre el número de proteínas y el número de cromosomas no siempre es de 1 a 1 (ver nota 4).

3. Los aminoácidos se pueden simbolizar con un código de tres letras o bien con un código de una letra (por ej., la lisina, Lys o bien K; la leucina, Leu o bien L, etc.)

4. La identidad de un gen, una proteína, es una simplificación. A veces, como en el caso de los anticuerpos, familias de genes codifican por familias de proteínas, combinando la información de los genes se traducen diferentes proteínas. En bacterias se puede dar el caso que una misma secuencia genética se empiece a transcribir a partir de diferentes puntos de la secuencia del gen, de manera que se acaben traduciendo proteínas diferentes.

genes que pueden estar activos en un determinado tejido. Hay genes del genoma que son activos en todos los tejidos y los hay específicos para cada tejido. El conjunto de proteínas que se sintetizan en una célula se denomina el «proteoma». El proteoma dependerá de la naturaleza del tejido a que pertenezca la célula y del estado fisiológico en que se encuentre la célula en aquel momento.

El proceso de activación de los genes del genoma empieza con el desarrollo embrionario. El proceso biológico que conduce desde el cigoto (la célula resultante de la fusión del óvulo con el espermatozoide) hasta el individuo adulto se explica en buena medida por la acción de las proteínas que se van sintetizando coordinadamente en distintas regiones del embrión. Cuando las células se han diferenciado en un tejido determinado pierden para siempre la facultad de producir unas determinadas proteínas. En este hecho se encuentra la base de la polémica sobre la clonación de embriones con fines médicos, porque es sólo en las primeras fases del desarrollo embrionario que las células mantienen todavía su plenipotencia, es decir, la posibilidad de producir cualquiera de las proteínas codificadas en el genoma. Las células de los embriones jóvenes tienen el potencial de transformarse en cualquier tejido. Este hecho hace de los embriones un material de mucho valor para la génesis *in vitro* de tejidos y para el desarrollo de nuevos agentes terapéuticos.

Entre el genoma de los humanos y el de los chimpancés hay una semejanza inquietante. Sólo un 1% de los caracteres del mensaje genético son diferentes (Calafell, 2000). Por tanto, inmediatamente pensamos que cuando sepamos qué proteínas codifica la parte del mensaje genético que nos hace diferentes de los chimpancés habremos dado un paso muy importante para conocer la *esencia* molecular humana; no obstante, hay estudios que indican que las diferencias obvias entre un chimpancé y un individuo de nuestra especie pueden deberse más al orden con que se activan los genes durante el desarrollo embrionario que no a las diferencias de la secuencia genómica (Jacob, 1998).

De manera que la capacidad de adaptar la célula a los posibles cambios fisiológicos y el «plan» de activación y desactivación de genes durante el desarrollo embrionario ha de computarse también como parte de la información genética del ADN.

La belleza del ADN se debe a un hecho tridimensional: es la clave para el estudio del metabolismo, para el estudio del desarrollo ontogenético y también para el estudio filogenético de los seres vivos. Acabamos de ver a qué se debe la variabilidad de las células de un individuo. Al final de esta lección hablaremos de las causas de la variabilidad entre los individuos de una misma especie. Permítanme explicar ahora otro tipo de información contenida en el mensaje genético.

Entre las diferentes especies, la información genética se diferencia por la longitud de la secuencia de bases y por el texto que contiene la secuencia. Los genomas más pequeños son los genomas de los virus, tienen entre tres mil y cien mil bases⁵; muy lejos de los tres mil millones de bases de los humanos. Si se consideran todas las especies existentes, se puede observar que la medida del genoma es mayor según se avanza en la complejidad de la especie. Las formas de vida unicelulares, como las bacterias, que ya disponen de una maquinaria celular para multiplicarse autónomamente,

5. El año 1977 se secuenció completamente el genoma del bacteriófago (X174, que tiene 5386 bases (F. Sanger *et al.*, 1977)

tienen genomas que oscilan entre medio millón y cincuenta millones de bases. A continuación, con genomas entre diez y cincuenta millones de bases encontramos las formas de vida pluricelulares, con poca comunicación intercelular, como los hongos. Finalmente están las plantas y los animales, que son formas de vida pluricelulares, con funciones celulares bien diferenciadas, es decir, con tejidos y muchos mensajeros moleculares. Las especies pertenecientes a este último grupo tienen genomas que oscilan entre los cincuenta y los miles de millones de bases. No obstante, a excepción de los mamíferos, pájaros y reptiles, dentro de cada grupo hay una gran variación de la medida del genoma, hasta el punto que algunas especies sobrepasan a otras que nos atreveríamos a calificar como más complejas. Es el caso de algunos anfibios que tienen un 60% más de ADN que la especie humana. Ya veremos, no obstante, que más ADN no significa necesariamente más genes y, por lo tanto, más proteínas.

Aparte de la medida, la estructura física del genoma puede ser muy diversa. Las células de las especies más evolucionadas, como las de los humanos, tienen el material genético duplicado⁶; el genoma de estas células se encuentra en forma de fragmentos lineales y confinado en un orgánulo celular, el núcleo. Cada uno de estos fragmentos se llama cromosoma. La especie humana tiene 23 parejas de cromosomas; en total, pues, 46. Los chimpancés tienen 48.

A pesar de las diferencias que acabamos de mencionar, hay muchas partes de la secuencia genética que se conservan con más o menos similitud en todas las especies, cosa que sólo se explica aceptando un determinado grado de parentesco entre todos los seres vivos.

Dos de las características más notables del ADN parece que se contrapongan. Por un lado, la molécula de ADN se puede replicar y transferir a la descendencia de forma prácticamente inalterada a lo largo de miles de generaciones, de manera que los humanos vamos teniendo hijos humanos y los monos, para poner un ejemplo, siguen teniendo hijos monos. Por otro lado, el DNA puede experimentar mutaciones de manera natural, es decir, el texto que contiene la secuencia de información genética cambia a lo largo del tiempo. En el caso más sencillo una mutación consiste en la sustitución de una base de la secuencia por otra base, no obstante, hay mutaciones que afectan a segmentos más o menos largos del genoma. Según hemos dicho anteriormente, si un gen experimenta un cambio, también tendrá que experimentarlo la proteína que codifica y esto puede afectar a la función de esta proteína. Hay que decir, con todo, que el código genético tiene unas características que permiten atenuar en buena medida el efecto de muchas mutaciones⁷.

La sustitución de un aminoácido por otro en una proteína como consecuencia de una mutación genética puede afectar la función de la proteína en grados diferentes. La magnitud de las repercusiones sobre el organismo dependerá de la función de esta proteína. Por ejemplo, la llamada anemia celular falciforme es una enfermedad debida a una mutación en único nucleótido del gen de la hemoglobina. Esta mutación provoca que los glóbulos rojos de los individuos afectados ten-

6. Excepto en el caso de las células germinales, que son haploides, es decir, con una única copia del material genético.

7. Todos los aminoácidos, excepto la metionina y el triptófano, están codificados por más de un codón (triplete de bases). Con las cuatro bases A, T, C i G se pueden hacer 64 combinaciones de tres bases; estas combinaciones se reparten entre los 20 aminoácidos. El código genético ha evolucionado de tal manera que una mutación en una de las bases comporta la traducción del mismo aminoácido o de otro aminoácido con propiedades fisicoquímicas parecidas.

gan forma de hoz y que se puedan ocasionar problemas circulatorios (Vella, 1980). En cambio, una mutación en una de las proteínas que intervienen en la formación de los cartílagos de la rata tiene un efecto múltiple en la anatomía de los animales afectados: aumento del volumen de las costillas, constricción del conducto traqueal, pérdida de elasticidad de los pulmones, hipertrofia del corazón, orificios nasales cerrados y morro chato (Strickberger, 1986). En muchos casos, no obstante, las consecuencias de una mutación son simplemente letales.

Es precisamente esta capacidad de experimentar cambios que tiene el ADN lo que permite comprender, en términos generales, la teoría de la evolución de las especies de Darwin. Como ya sabemos, esta teoría explica cómo los cambios aleatorios que se producen en un único individuo pueden llegar a cambiar, en algunas generaciones, el mensaje genético de una especie.

Uno de los descubrimientos más útiles y a la vez más sorprendentes para el estudio de la evolución de las especies ha sido el llamado «reloj molecular» (Cavalli-Sforza, 1997). Los genes, como acabamos de decir, experimentan mutaciones y estas mutaciones se pueden ver reflejadas en las proteínas que codifican. Pues bien, se ha podido observar que en especies para las cuales se conoce el tiempo de separación en el árbol filogenético gracias a métodos paleontológicos, el número de aminoácidos diferentes en una misma proteína de las dos especies se puede correlacionar muy bien con el tiempo de separación. Esto significa que la tasa de cambios para una proteína determinada es constante. De manera que contabilizando los cambios de esta proteína en dos especies actuales se puede estimar cuanto tiempo hace que empezó a divergir, dejando atrás su ancestro común. Utilizando este método se pudo establecer que la línea del árbol filogenético que lleva hasta los humanos se separó de la línea que lleva a nuestro pariente más cercano, el chimpancé, hace cinco millones de años⁸ (Ruvolo, 1997).

Nuestra especie, el *Homo sapiens*, tiene una historia que empezó muy posiblemente en África hace aproximadamente ciento cincuenta mil años. Los estudios de variabilidad genética que dieron soporte al origen africano de los humanos se llevaron a cabo con un material genético un poco especial: el ADN mitocondrial. Las mitocondrias son unos orgánulos celulares que disponen de una dotación de ADN diferente de la del núcleo celular. Se cree que estos orgánulos provienen de una antigua simbiosis entre las bacterias y las células de los organismos superiores. Esto se habría producido hace un millón de años. De hecho, entre el ADN de las mitocondrias y el ADN de algunas bacterias actuales existen muchas similitudes. En cada célula existen miles de mitocondrias. La función de estos orgánulos es producir energía química útil para las reacciones metabólicas. En la fecundación, las mitocondrias del espermatozoide no se incorporan al cigoto, se pierden; de mane-

8. La formación de una nueva especie a partir de una especie existente requiere como primera condición el aislamiento sexual de una población de la especie del resto de individuos, por ejemplo, a causa de una migración. Cuando las dos poblaciones han acumulado independientemente un número crítico de cambios en el genoma a causa de mutaciones, recombinación genética y selección natural, se pierde la capacidad de reproducción entre los individuos de estas dos poblaciones. No se trata de un impedimento morfológico sino de un bloqueo de la reproducción a nivel molecular. Mientras el cruce entre los individuos de una población sea posible los cambios individuales se repartirán aleatoriamente a lo largo de las generaciones sin que se puedan establecer dos líneas evolutivas diferentes dentro de la población. Ciertamente, esta situación sería materialmente posible, pero la probabilidad de que se dé es muy reducida.

ra que las mitocondrias del nuevo ser son descendientes de las mitocondrias presentes en el óvulo materno. El ADN de estos orgánulos se transfiere a la descendencia por vía materna. Esta característica⁹ del ADN mitocondrial facilita la construcción de genealogías, ya que no es la mezcla del material genético de los progenitores. Con el análisis del ADN mitocondrial completo de un japonés, un europeo y un africano y siguiendo el razonamiento que más variación en el ADN significa más tiempo de separación, se pudo comprobar que el ADN mitocondrial africano era el más antiguo (Horal *et al.* 1995). De aquí que la revista de moda *Vogue*, solicitase una entrevista a L.L. Cavalli-Sforza (el prestigioso geneticista de poblaciones) para que explicase cuanto tiempo hacía que había nacido la *Eva africana* (Cavalli-Sforza, 1997). La construcción precisa del árbol filogenético de las poblaciones humanas actuales requiere, no obstante, el estudio a fondo de la variabilidad del ADN de los cromosomas del núcleo. Estos ejemplos ilustran cómo el mensaje genético incluye también información sobre la historia evolutiva de los organismos y cómo el análisis de las diferencias entre las especies a nivel molecular puede permitir la reconstrucción del árbol genealógico de los seres vivos.

Ahora hacemos un pequeño cálculo¹⁰. Los seres humanos tenemos aproximadamente cien mil proteínas. Según el código genético, cada aminoácido se identifica en la secuencia del ADN mediante tres letras del alfabeto, C, T, G y A¹¹. El número de aminoácidos de una proteína típica es de aproximadamente cuatrocientos. Una proteína de esta medida requiere, pues, una secuencia de ADN de mil doscientas bases. Para cada proteína hay que añadir quinientas bases más que corresponderían a secuencias reguladoras y a los llamados «intrones»¹². Las cien mil proteínas requieren, por tanto, unos doscientos setenta millones de caracteres de texto genético. Puesto que el genoma humano tiene una medida de 3340 millones de bases, tiene que haber aproximadamente un 95% del ADN sin ningún gen¹³. Las mutaciones de que hemos estado hablando no sólo se dan en los genes, de manera que entre los individuos de una misma especie también habrá diferencias genéticas en las zonas no codificadoras de su ADN.

Por polimorfismo hemos de entender cualquier diferencia que se pueda detectar en la secuencia genética de los individuos de una misma especie. A lo largo del genoma humano, por ejemplo, hay unos cien mil segmentos de ADN constituidos por repeticiones de la pareja de bases CA. Estos bloques se denominan microsatélites. Los microsatélites pueden tener de una a cuarenta repeticiones seguidas de la pareja de bases CA y se transmiten a la descendencia conservando la medida del bloque. El polimorfismo consiste en cambios de la medida de un determinado microsatélite en diferentes individuos de la población. Otros polimorfismos importantes, que se encuentran tanto en los genes como en el resto del ADN, son los llamados «polimorfismos de dianas de restricción» (RFPL), que no comentaremos, y los «snips», entre otros. «Snip» es la pronunciación inglesa

9. Además, en el ADN mitocondrial no hay recombinación genética y la tasa de mutación es mayor.

10. Este cálculo sólo pretende ser ilustrativo.

11. Ver la nota 7.

12. Los intrones son secuencias del ADN de un gen que no se encuentran en el RNA mensajero maduro.

13. Este ADN que no codifica ninguna proteína se considera ADN estructural. Parece ser que es necesario para que los genes se puedan transcribir a más velocidad (Baldi y Brunak, 1998).

de las siglas SNP, que traducidas significan «polimorfismo nuclear único». Los SNP son diferencias en una única base, en alguna posición del mensaje genético, de los individuos de la población. Actualmente se está elaborando una base de datos en la cual se introducen los SNP que se van descubriendo (<http://snp.cshl.org>). Se prevé que en la primavera del 2001 se habrá construido un mapa del genoma humano con la posición y naturaleza de aproximadamente setecientos mil polimorfismos nucleares únicos (Roses, 2000). Una de las ventajas de los SNP es que se está desarrollando la tecnología, los llamados «chips de ADN»¹⁴, que ha de permitir llevar a cabo el análisis simultáneo de miles de SNP automáticamente.

Hemos visto que las células de los humanos tienen una dotación doble de material genético, por tanto, cada individuo puede tener como máximo dos versiones de un determinado polimorfismo, una versión situada en cada uno de los dos cromosomas homólogos. Las distintas versiones de un gen se denominan clásicamente alelos, este concepto se ha hecho extensivo a todos los polimorfismos, se encuentren o no en una secuencia codificadora. Un ejemplo familiar para todo el mundo son los alelos que determinan el grupo sanguíneo de los humanos.

En el proceso que lleva a la formación de los gametos masculino y femenino, que sólo tienen una copia de material genético, es decir, en que los cromosomas no están duplicados, se producen dos fenómenos que son la causa principal de la diversidad de los individuos de una población. Por un lado, los 23 cromosomas de un óvulo o un espermatozoide pueden provenir de cualquiera de los cromosomas de la pareja del individuo adulto. Esto hace que se puedan producir más de ocho millones (2^{23}) de tipos diferentes de gametos en relación a su dotación cromosómica. Por otro lado, a lo largo del proceso de formación de los gametos, mediante un fenómeno llamado «recombinación genética», los cromosomas homólogos se intercambian material genético, de manera que los diferentes alelos se redistribuyen a lo largo del cromosoma, cosa que hace aumentar aún más el número posible de gametos genéticamente diferentes. Esta es la causa por la cual los hermanos de una familia son diferentes; los hijos reciben una combinación aleatoria de las características de los padres.

Gracias al impulso del Proyecto Genoma Humano¹⁵, en la base de datos GenBank¹⁶, en los EEUU, se entran 1.700.000 nuevas bases por día pertenecientes a secuencias genéticas de representantes de todos los grupos de especies (Baldi y Brunak, 1998). Estas secuencias están al alcance de todo el mundo a través de Internet de forma gratuita. En esta misma base de datos se puede encontrar la secuencia completa de diversos organismos, entre los cuales *Escherichia coli* (una bacteria), *Saccharomyces cerevisiae* (una levadura), *Arabidopsis thaliana* (una planta), una buena parte de la secuencia de *Drosophila melanogaster* (un animal), los cromosomas 21 y 22 humanos y miles de fragmentos del mensaje genético de centenares de especies. Parece claro que el estudio de la filogenética de los seres vivos mediante el análisis de la diversidad genética de las especies actuales

14. Se trata de unos soportes, de aproximadamente un centímetro cuadrado, en los que se inmovilizan las sondas de ADN adecuadas para detectar un determinado conjunto de SNP.

15. Con la tecnología del año 1993 se podían secuenciar 90 millones de bases al año a un coste de 100 pesetas por base; actualmente se secuencian 500 millones de bases al año, a un precio de 50 pesetas por base.

16. <http://www.ncbi.nlm.nih.gov>

tiene un gran futuro. Y aquí cabe añadir que la tecnología para analizar el ADN de organismos fósiles pronto será una práctica al alcance de todos los laboratorios, cosa que permitirá afinar, quién sabe si incluso cambiar, el conocimiento actual sobre la historia de la evolución.

La posibilidad de determinar con métodos fiables y rápidos (como en el caso de los SNP) la diversidad del genoma de cada individuo, además del interés puramente científico que como hemos visto supone, ha de permitir diagnosticar y tratar con más eficacia muchas de las enfermedades que por ahora escapan a los tratamientos farmacológicos. La farmacogenética es el estudio de cómo las diferencias genéticas condicionan la respuesta de los pacientes a un determinado medicamento. Virtualmente, toda enfermedad de carácter no infeccioso lleva implícita la disfunción de una proteína. Con frecuencia los fármacos se diseñan para actuar sobre estas proteínas. Ahora bien, pequeñas diferencias en una proteína, debidas a la variabilidad genética del individuo, pueden atenuar e incluso invalidar los efectos de un fármaco. O, visto desde otro ángulo, pueden provocar unos efectos secundarios más perniciosos en unos individuos que en otros (Roses, 2000). De manera que es posible concebir (y de hecho las compañías farmacéuticas ya trabajan en ello) fármacos para una determinada enfermedad, específicos para el perfil genético del enfermo.

Pero aparte de todo lo que acabamos de decir, la ingente cantidad de datos generales para el Proyecto Genoma Humano empieza a cambiar la manera de estudiar el fenómeno vital.

Los organismos son redes de genes que hacen redes de proteínas que regulan los genes¹⁷. Por tanto, vemos que otra de las funciones de las proteínas que no hemos mencionado al principio de la lección es precisamente interactuar físicamente con el ADN para controlar la síntesis de proteínas. Es un pez que se muerde la cola. Quién controla pues al sistema de control? La respuesta es que el sistema se autocontrola. Esta conclusión no sorprenderá en absoluto a los ingenieros, acostumbrados a tratar con mecanismos de este tipo. Ahora bien, si se quiere tratar el funcionamiento molecular de un ser vivo como un sistema dinámico hay que conocer previamente cuales son los elementos de que está formado y cómo están relacionados entre ellos. El Proyecto Genoma Humano abre la posibilidad de iniciar este estudio. Se trata de pasar de un enfoque reduccionista de la biología a un enfoque holístico. Para esto hay que desarrollar las herramientas conceptuales, matemáticas e informáticas que permitirán detectar los genes en el desierto que representa la secuencia genómica, deducir, a partir de la secuencia del gen, la función de la proteína que codifica (de esto se empieza a ocupar un nuevo campo de la ciencia llamado «genómica funcional»), detectar entonces

17. La información contenida en el ADN no se traduce directamente a una secuencia de aminoácidos, sino que lo hace a través de la síntesis de un intermediario muy parecido en todos los aspectos al ADN, el llamado RNA, ácido ribonucleico. De manera que la información para la síntesis de proteínas fluye en el sentido DNA, RNA, proteínas. El proceso inverso no se conoce. Si bien algunos virus, los llamados retrovirus (entre los cuales se encuentra el virus del SIDA) pueden producir una molécula de DNA a partir de RNA, no se conoce ningún proceso bioquímico capaz de producir una molécula de un ácido nucleico a partir de la información contenida en una proteína. Cuando se intenta comprender el origen de la vida conforme a una evolución química de tipo darwiniano chocamos con una paradoja del tipo "qué fue primero, el huevo o la gallina?": la información para la síntesis de las proteínas se encuentra contenida en el ADN, y el ADN no puede sintetizar proteínas sin la intervención, precisamente, de las proteínas. Una posible solución a la paradoja pasa por el descubrimiento de las ribozimas: moléculas de RNA con actividad enzimática (Schuster, 1995).

las proteínas que interactúan para llevar a cabo una determinada función y establecer finalmente la dinámica de esta red. Trabajo de sobra para el siglo XXI.

Llegados a este punto de la lección, la pregunta que podemos hacernos es si el conocimiento completo del mensaje que contiene el ADN nos permitirá descubrir alguna cosa esencial de la naturaleza humana. Yo no me atrevería a pronunciarlo. No obstante, me gustaría poner dos ejemplos de proteínas relacionadas con fenómenos mentales. En la mosca del vinagre, *Drosophila melanogaster*, se han podido aislar dos proteínas que intervienen en la memoria a largo término, la una activa este tipo de memoria y la otra la inhibe. Las moscas transgénicas, diseñadas para que sintetizen una cantidad mayor de la proteína activadora, sólo necesitan una sesión de entrenamiento para aprender a asociar un determinado olor con una corriente eléctrica; las moscas normales requieren dos sesiones (Connolly y Tully, 1996; Tully, 1998)¹⁸. El segundo ejemplo está relacionado con la conciencia. En la transmisión de señales a lo largo del sistema nervioso intervienen unas estructuras proteínicas, los microtúbulos, que forman el llamado esqueleto celular o citoesqueleto. La pérdida de conciencia causada por anestésicos como el éter o el cloroformo es debida a un efecto sobre la tubulina, la proteína que forma los microtúbulos. La tubulina se puede encontrar en dos formas tridimensionales diferentes, que si queremos podemos simbolizar con un 0 y un 1. Aunque no se dispone de pruebas definitivas, parece ser que el anestésico bloquea la «conmutación» entre las dos formas; la tubulina no podría pasar de 0 a 1 o de 1 a 0. Desde un punto de vista matemático, los microtúbulos tienen una estructura muy interesante: son cilindros vacíos constituídos por 13 columnas de moléculas de tubulina. Si nos imaginamos el cilindro desplegado en un plano, las moléculas de tubulina se ordenan en líneas inclinadas que llegan del centro a los extremos opuestos con un desplazamiento de 5 u 8 lugares. El 5, el 8 y el 13 son tres números consecutivos de la serie de Fibonacci. La disposición de las moléculas de tubulina conforme a números de Fibonacci puede favorecer, según dicen los matemáticos, la función del microtúbulo como «procesador de información». Los microtúbulos y la tubulina podrían llevar a cabo computaciones al estilo de un «autómata celular» y transportar los mensajes a lo largo del sistema nervioso (R. Penrose, 1996).

Me gustaría acabar esta lección con unas palabras de Umberto Eco, extraídas de su libro *Sis passejades pels boscos de la ficció*. Dice lo siguiente: «Hay una regla áurea para todo criptoanalista o descriptor de códigos secretos, y es que todo mensaje puede ser descrito siempre que se sepa que se trata de un mensaje. El problema con el mundo real es que nos estamos preguntando, desde hace milenios, si hay un mensaje y si este mensaje tiene un sentido».

Muchas gracias.

18. Ejemplo extraído de «La comprensión de la conducta: aportaciones de la genética molecular» de Rosa M. Escoriuela, en *La biología a l'alba d'un nou mil·lenni*. Treballs de la Societat Catalana de Biologia, Volum 50. Barcelona: Jaume Bertranpetit Editor, 2000.

Bibliografia

- Baldi, P. y Brunak, S. *Bioinformatics*. Massachusetts: Bradford Book, 1998.
- Calafell, F. La diversitat genètica humana: de l'espècie a l'individu, dins *La Biologia a l'alba d'un nou mil·lenni*. Treballs de la Societat Catalana de Biologia, volum 50. Barcelona: Jaume Bertranpetit Editor, 2000.
- Cavalli-Sforza, L.L. *Gens, pobles i llengües*. Barcelona: Proa, 1997
- Connolly, J.B. y Tully, T. You must remember this. *The Sciences*, mayo/junio, 37-42 (1996).
- Dunham, I. *et al.* The DNA sequence of human chromosome 22, *Nature*, 402, 2 diciembre, 489-495 (1999).
- Eco, H. *Sis passejades pels boscos de la ficció*. Barcelona: Destino, 1997.
- Hattori, M. *et al.* The DNA sequence of human chromosome 21, *Nature*, 405, 18 mayo, 311-319 (2000).
- Holliday, R. The history of DNA heteroduplex. *BioEssays*, 12, 133-142 (1990).
- Horal, S. *et al.* Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc. Natl. Acad. Sci.*, 92, 523-536 (1995).
- Jacob, F. *La mosca, el ratolí i l'home*. Barcelona: Edicions 62, 1998.
- Penrose, R. *Las sombras de la mente*. Barcelona: Grijalbo Mondadori, 1996.
- Roses, A.D. Pharmacogenetics and the practice of medicine. *Nature Insight*, 405, 857-865 (2000).
- Ruvolo, M. Molecular phylogeny of the hominoids: inferences from multiple independent DNA sequence data sets. *Mol. Biol. Evol.*, 14, 248-265 (1997).
- Sanger, F. *et al.* Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*, 256, 687-695 (1977).
- Schuster, P. Minimal properties for evolutionary optimization, dins *Cooperation and conflict in general evolutionary processes*. New York: John Wiley and Sons, Inc. 1995.
- Strickberger, M.W. *Genética*. Barcelona: Omega, 1986.
- Vella, F. Human haemoglobin and molecular disease. *Biochem. Educ.*, 8, 41-53 (1980).