

BIOLOGÍA MOLECULAR: LA NUEVA FRONTERA

Por EDUARDO CADENAS

UNIVERSIDAD DE ALICANTE

MCMLXXXVI

Discurso Inaugural

Apertura del curso 1986-1987

“Así que conociendo cuál es la forma de los cuerpos inanimados o minerales estaremos tanto mejor capacitados para seguir nuestra búsqueda siguiente acerca de las formas de los cuerpos vegetales y por fin, a la de los animados, que parece la etapa superior del conocimiento natural del que es capaz la mente del hombre”

Robert Hooke, MICROGRAPHIA, 1665.

“Un aspecto central de la biología molecular es la comunicación dinámica entre los elementos genéticos, representados por los ácidos nucleicos, y el aparato metabólico, representado por las enzimas y otras proteínas”.

H. J. Vogel, 1962

Molt Honorable Senyor President de la Generalitat  
Excelentísim y Magnífico Señor Rector  
Excelentísimos e Ilustrísimos Señores  
Señoras y Señores

Siguiendo una vieja tradición universitaria hoy me ha correspondido el honor de dictar esta lección inaugural. El privilegio está basado en la antigüedad. Constituye una reliquia de otras épocas y otras maneras de entender las relaciones humanas. Pocos vestigios así consiguen sobrevivir y es posible que éste también sea sometido a revisión. También ha sido tradicional que la lección inaugural fuese la lección de introducción a una disciplina concreta. En esta ocasión hemos escogido como tema hablaros de la Biología Molecular, su historia, su situación actual de frenético desarrollo y las perspectivas que nos presenta para los próximos años, en los que bien pudiera tener lugar un segundo renacimiento como consecuencia de la aplicación de la Biología Molecular al Hombre.

### ORIGEN DE LA BIOLOGÍA MOLECULAR

Parece que el término de Biología Molecular fue acuñado por W. Weaver de la Rockefeller Foundation en 1938. Estaba preparando un plan de apoyo a la investigación para la aplicación de la ciencia física a áreas seleccionadas de la Biología, como son la Bioquímica, la Biología Celular y la Genética. El programa de la Fundación tuvo un éxito espectacular. Ayudó entre otros, a investigadores tan destacados como Rudolf Schoenheimer, Linus Pauling, Robert Robinson y George Beadle. Pero el término Biología Molecular no se consolidó. Quizá fuese porque pudo parecer demasiado presuntuoso, pero más probablemente se debió a la interrupción del estudio de las macromoléculas, que supuso el inicio de la Segunda Guerra Mundial.

Fue después a la terminación de la contienda, cuando se popularizó el término y se llegó a perfilar su contenido. Desde el primer momento surgieron dos escuelas que se disputaron la hegemonía. G. S. Stent las ha descrito como escuela informacionista y escuela estructuralista. La primera, americana, era hostil a la Bioquímica, la segunda, inglesa, estaba en cambio, plenamente integrada en ella. Las dos en estrecha conexión con la Física, pero entendiendo la relación de formas contrapuestas. Algunos pioneros de la escuela informacionista creían en la idea realmente fantástica de que la Biología podía proporcionar contribuciones significativas al progreso de la Física, incluso nuevas leyes y fenómenos. Los pioneros de la escuela estructuralista mantenían el punto de vista perfectamente razonable, de que la Física podía hacer aportaciones muy valiosas a la Biología. (Stent, 1968).

## ESCUELA INFORMACIONISTA

La escuela informacionista de la Biología Molecular tuvo su antecedente en las ideas primero de Niels Bohr y después de su discípulo Max Delbrück. Cuando el vitalismo pasado de moda estaba desapareciendo de los círculos intelectuales, Bohr elaboró la noción de que algunos fenómenos biológicos no se podrían explicar completamente en términos de la Física convencional. En opinión de Bohr la dificultad de entender la vida en esos términos radica en que “*las condiciones que prevalecen en la investigación biológica y física no son directamente comparables, puesto que la necesidad de mantener al objeto de investigación vivo impone restricciones en la primera, para las que no hay contrapartida en la segunda. Así parece existir en el animal vivo un “principio de incertidumbre” semejante al del electrón*”. En 1935 Max Delbrück afirmó que la genética era el dominio de la investigación biológica en que las explicaciones físicas y químicas podían resultar insuficientes en el sentido expresado por Bohr, de modo que condujo que: “*la genética es autónoma y no se debe mezclar con concepciones fisicoquímicas*”.

El interés de los físicos en la Biología parece que, en gran medida, se debió a la influencia de uno de los líderes de la física teórica. En 1945 apareció un libro pequeño titulado “What is life?” escrito por Edwin Schrödinger en que se anunciaba a los físicos una nueva era en la investigación biológica que había de proporcionar descubrimientos extraordinarios. Levantando las pasiones de su público, el libro se convirtió en una especie de “Uncle Tom’s cabin” de la revolución biológica. El libro tuvo mucha influencia debido a su elegancia de estilo y su calidad intelectual. Como decía Crick: “*el librito de Schrödinger fue muy oportuno y atrajo mucha gente a la Biología que, de otra forma, no se hubiera dedicado a ella en absoluto*”. (Crick, 1965). Y además centró la atención sobre el tema del material genético. Según Schrödinger del concepto general de Delbrück de la sustancia hereditaria se deduce que “*la materia viva aunque no elude las leyes de la Física ya establecidas, es probable que implique “otras leyes de la Física” que, no obstante, una vez descubiertas, pasarán a formar parte integral de esa ciencia como las otras*”. El libro lo leyó también Watson y después de haberlo hecho quedó polarizado hacia el objetivo de desentrañar el secreto del gen (Watson, 1966).

El estudio de los ácidos nucleicos descubiertos el siglo pasado se vio impulsado con fuerza en dos ocasiones, como consecuencia del conocimiento de su función biológica. Recién descubiertos se localizaron en el núcleo celular (de ahí su nombre) . Como consecuencia del trabajo de los citólogos se concluyó que el material genético tenía su sede en este orgánulo celular. De modo que para la década de los ochenta del pasado siglo ya estaba claro que el DNA debía ser el material genético. Como se sabe sería precisamente el progreso en el conocimiento de su estructura, realizado en condiciones deficientes, lo que llevaría a continuación al abandono de tal noción, hasta los años 40 de nuestro siglo que fue cuando se demostró definitivamente que el DNA es el material genético. Los experimentos cruciales fueron los de Avery, MacCarthy y MacLeod, que probaron que el factor de transformación es DNA y los de Hershey y Chase que mostraron que el material genético del fago T2 es

su DNA. Si el DNA era el material genético debía contener información y no podía tener la estructura que la hipótesis tetranucleotídica la había asignado. E. Chargaff y su grupo realizaron una extensa determinación de la composición de DNA de las fuentes biológicas más variadas y efectivamente mostraron variación en la composición dependiendo del origen de la muestra. Además encontraron una curiosa regularidad en la composición. El contenido en adenina es igual al contenido en timina, y el contenido de guanina, igual al de citosina. (Chargaff, 1950).

Más decisivo fue aún el esfuerzo realizado por Watson y Crick quienes convencidos de que el DNA es el material genético se propusieron establecer su estructura, con la ilusión de que quizá la estructura una vez conocida, pudiera revelar como esta sustancia ejerce las diversas funciones que corresponden al material genético celular. (Watson y Crick, 1953).

Partieron de los datos de difracción de rayos X de fibras de DNA obtenidos por R. Franklin y M. Wilkins del King's College en el Reino Unido y trabajaron con modelos moleculares, como lo había hecho L. Pauling del California Institute of Technology para establecer los elementos de estructura secundaria de las proteínas fibrosas. En 1953 llegaron finalmente a la celebrada estructura de la doble hélice, en que el emparejamiento de bases complementarias por medio de enlaces de hidrógeno explicaba las regularidades en la composición observadas por Chargaff. La estructura efectivamente sugería como lleva a cabo una de las misiones asignadas al material genético: la transmisión fiel de la información de generación en generación. La complementariedad de las bases en ambas cadenas de la doble hélice les llevó a proponer que en la replicación del DNA cada cadena debe servir de modelo o molde para la síntesis de una nueva, con lo que de una doble cadena se pasa a dos idénticas entre sí e idénticas a la primera. También constituye la estructura del DNA el fundamento de las demás funciones biológicas estudiadas por la Genética. Como dijo Beadle en 1969: *"resolver la estructura del DNA es uno de los grandes hallazgos de la Biología del siglo XX, comparable en importancia a los logros de Darwin y Mendel en el siglo XIX... porque la estructura de Watson y Crick sugiere de inmediato cómo se replica o copia en cada generación celular, cómo se utiliza en el desarrollo y la función y cómo sufre cambios mutacionales, que son la base de la evolución orgánica"*.

## ESCUELA ESTRUCTURALISTA

Antes de que se lograse probar que el DNA es el material genético las proteínas habían sido el candidato generalmente aceptado. Por su importancia y variedad y la diversidad de las funciones que ejercen en la célula.

La presencia universal de la proteína en los seres vivos llevó a Johannes Mulder a proponer el siglo pasado el nombre que hoy tienen, queriendo indicar que son de la mayor importancia. Pero su estudio adquirió un ímpetu arrollador solo después de que tras larga polémica se logró demostrar que las enzimas son proteínas. Desde entonces profundizar en el conocimiento de la estructura

de las proteínas era lo mismo que ahondar en la comprensión de cómo ejercen su función estos catalizadores biológicos, de los que dependen todas las actividades vitales.

Durante décadas la Bioquímica se ocupó del estudio de las transformaciones químicas que ocurren en las células, descomponiendo la marcha del proceso en etapas individuales mediante el análisis de la estructura de los sustratos y productos. Entre tanto las enzimas que hacen posibles esas transformaciones no se estudiaban sino por medio de técnicas indirectas y sus propiedades sólo se lograban establecer por medio de inferencias. Pero llegó un punto en que el énfasis pasó al estudio de la estructura detallada de las proteínas y fue entonces cuando nació una nueva era que se denomina molecular, pero que estrictamente se debería llamar macromolecular, porque las proteínas son macromoleculares.

La escuela estructuralista de la Biología Molecular tiene sus orígenes en los estudios por difracción de rayos X de proteínas que se presentan en forma de fibras en que hay suficiente regularidad. Como decía W. S. Astbury: “*las fibras visibles o invisibles son los componentes estructurales principales de todos los tejidos biológicos... las fibras están formadas por moléculas en forma de cadenas y la biología molecular de momento se ocupa del partido que la vida ha sabido sacar de tales moléculas*”. Astbury comenzó su vida profesional como físico, pasó a la cristalografía de rayos X y por fin a la Biología Molecular. Su entrada en esta ciencia estuvo marcada por la impresión inolvidable que le causó la regularidad biológica revelada por la difracción de rayos X de las fibras textiles tradicionales, que son las fibras biológicas.

En 1950 Astbury decía: “*parece que el término Biología Molecular se está popularizando bastante y estoy contento de que así sea, porque, aunque es improbable que haya sido yo quien la ha inventado, me gusta y he tratado de propagarlo durante mucho tiempo*”. (Astbury, 1950) . A continuación señalaba lo que caracteriza esta forma de ver la Biología: “*implica no tanto una técnica, sino más bien un nuevo enfoque desde el punto de vista de las llamadas ciencias básicas, con la intención de buscar bajo las manifestaciones a gran escala de la Biología clásica, el plan molecular correspondiente*”. Por fin señalaba claramente el objeto de estudio: “*se ocupa particularmente de las formas de las moléculas biológicas y de la evolución, explotación y ramificaciones de esas formas en su ascenso hacia niveles de organización cada vez más altos*”. Para terminar definiéndola: “*la Biología Molecular es predominantemente tridimensional y estructural, pero esto no quiere decir, sin embargo, que sea un mero refinamiento de la morfología. Al mismo tiempo tiene que inquirir forzosamente sobre génesis y función*”.

Sin duda alguna es ésta una excelente definición de la Biología Molecular para la escuela estructuralista cuyos antecedentes son las técnicas físicas y su aplicación a temas químicos, en particular la difracción de rayos X aplicada a la elucidación de la estructura de moléculas de trascendencia biológica. W. H. Bragg y su hijo W. L. Bragg inventaron la cristalografía de rayos X en 1912 y luego fundaron la escuela de cristalograjos que convirtió al Reino Unido en la patria de la estructura molecular. El éxito de la técnica

condujo a su aplicación a moléculas cada vez más complicadas. Fueron discípulos suyos, como Atsbury y J. D. Bernal, los que en los años 30 comenzaron a investigar la estructura de las proteínas y los ácidos nucleicos convencidos de que la función fisiológica de la célula se puede llegar a comprender sólo en términos de la estructura tridimensional de sus componentes. Algunos de sus trabajos proporcionaron los primeros atisbos del campo en que luego habría avances espectaculares. Así Bernal consiguió demostrar en 1939 que el virus del mosaico del tabaco, TMV, consiste en una asociación de centenares de subunidades proteínicas idénticas y Astbury en 1945 que en el DNA las bases forman una pila compacta, que es perpendicular al eje de la molécula.

Fue a comienzos de la década de los 30 cuando Astbury consiguió establecer la existencia de dos patrones característicos en las figuras de difracción de las proteínas fibrosas, que denominó alfa y beta. No obstante, la interpretación no ocurrió hasta los años 50 y no la propuso un miembro de la escuela británica, sino Pauling en California. Utilizó modelos moleculares desarrollados y construidos por él e hizo uso del concepto de enlace de hidrógeno, que también él había desarrollado. Así pudo proponer la existencia de los elementos de estructura secundaria, denominados hélice alfa y hoja beta, que daban cuenta de las pautas de difracción descritas por Astbury.

La interpretación de la estructura que realizó Pauling indicaba una disposición periódica de la cadena polipeptídica para generar un entramado extremadamente ordenado. La visión aunque extraordinaria bajo el punto de vista estético, supuso una gran decepción, por cuanto no sugería nada acerca de cómo se sintetizan las proteínas, ni sobre cómo logran desempeñar las variadas funciones biológicas que la célula les ha confiado. (Pauling et al., 1951).

Es curioso que los primeros avances espectaculares en nuestro conocimiento sobre la estructura biológica tuviesen su arranque en el estudio de las proteínas fibrosas. La razón de que utilizase este tipo de material fue doble. Por un lado la facilidad de su obtención en cantidad sin más guía que sus especiales propiedades fisicoquímicas; por otro, la regularidad de la estructura de las moléculas en sí mismas y de su disposición en la fibra. Pero las proteínas fibrosas son una originalidad evolutiva de las células eucarióticas que sin duda tiene mucho que ver con su habilidad para constituir organismos pluricelulares. Se trataba por tanto de proteínas poco frecuentes. La mayoría de las proteínas son globulares, actúan en disolución acuosa y están formadas por la agregación de subunidades.

En 1937 Max Perutz se propuso en Cambridge establecer la estructura en el espacio de la molécula de la proteína hemoglobina, por difracción de rayos X. En este caso no se contaba con fibras, sino con cristales y el propósito no era explicar determinadas regularidades de la estructura, sino obtener la verdadera disposición, completamente desconocida, de los átomos en una molécula formada por la agregación de cuatro subunidades, cada una provista de un grupo hemo. Se había escogido la hemoglobina por ser fácil de obtener y ser relativamente pequeña, así como por la enorme importancia fisiológica de

su función de transporte de oxígeno en la sangre y la gran riqueza de estudios fisicoquímicos realizados sobre ella. Parecía probable que la estructura debería iluminar la función en este caso, como en ningún otro. En 1959 M. Perutz junto con M. Rossmann, A. F. Cullis, H. Muirhead y T. C. T. North lograban resolver la estructura tras 22 años de esfuerzo. Cuando se acometió la empresa era imposible de realizar, pero al paso de los años se resolvieron los problemas teóricos y prácticos haciendo posible coronarla con éxito. Antes que Perutz, J. Kendrew lograba en 1957 revelar la estructura de otra proteína globular más simple que la hemoglobina, la mioglobina que en el músculo, especialmente de mamíferos marinos, almacena oxígeno para suplir sus necesidades respiratorias. (Dickerson y Geis, 1981).

Quizá al aspecto más curioso de ambas estructuras fuese el que incluyera numerosas hélices alfa con una estructura exactamente coincidente con la que Pauling había propuesto para las proteínas fibrosas, basándose en sus modelos moleculares. Pero las hélices no estaban dispuestas de ninguna forma en una estructura regular y la impresión que causaba la molécula en palabras de Perutz era visceral y repugnante. Una vez más la estructura resultaba decepcionante por cuanto no parecía iluminar cómo la hemoglobina ejerce su función. La cooperatividad en la unión con el oxígeno, conocida desde comienzos de siglo y atribuida a una interacción directa entre los grupos hemo, quedaba descartada por el alejamiento de estos grupos en la molécula.

Coincidiendo en el tiempo en buena medida con el proyecto de determinación de la estructura de las primeras proteínas globulares, se fue convirtiendo en realidad otro intento brillante. Fue el trabajo de Frederick Sanger y su grupo que se proponía la secuenciación de una proteína. Cuando lo acometió parecía razonable suponer que todas las moléculas de una proteína tuvieran los mismos residuos de aminoácido y que estos residuos estuviesen dispuestos en el mismo orden. Pero este punto de vista, que sostenía Sanger, no era compartido por todos los investigadores, muchos de los cuales pensaban que, de ser cierto, supondría una enorme carga para la célula. Sanger realizó un trabajo enormemente original. Tuvo que preparar la estrategia de la secuenciación, poner a punto las técnicas necesarias, algunas de las cuales tuvo que desarrollar él mismo y por fin realizar su propósito mediante un esfuerzo continuado de casi una década. Comenzó el proyecto en 1944. Eligió la hormona insulina porque era relativamente pequeña: solo tiene 51 residuos y se disponía de ella en cantidad. Además se creía que el conocimiento de su secuencia quizás revelase cómo actúa para provocar sus numerosas acciones celulares, que contribuyen eficazmente a mantener el nivel de glucosa en sangre, dentro de límites apropiados para el organismo animal. Consiguió desarrollar un método muy conveniente para marcar residuos N terminales, que fue decisivo para el éxito del proyecto. Finalmente en 1953 Sanger y Thomson ya habían completado la secuencia de las cadenas A y B de la hormona. Dos años después lograron establecer la posición de los puentes disulfuro en la molécula de la insulina. Así habían probado que todas las proteínas tienen secuencia y por ello contienen información y también acababan de probar por primera vez que una cadena polipeptídica no contiene otro tipo de enlaces que los enlaces peptídicos.

Una vez desarrolladas las técnicas que permitían secuenciar una proteína y establecer su estructura en el espacio, su aplicación permitió conocer en detalle muchas otras. Pero al principio el conocimiento de la estructura no abrió las puertas a la comprensión de cómo las proteínas ejercen las acciones que las caracterizan. Pronto se propuso que le problema estribaba en que la molécula proteínica es muy grande y en buena parte, la estructura no es relevante a la acción. Sólo debe serlo una parte que es preciso identificar. Con el propósito de contar con los criterios que lo hiciesen posible se inició una fértil área de investigación, en que a través de la comparación de la misma proteína en especies diferentes, se esperaba encontrar las regiones conservadas de la molécula y en ellas la base de la acción biológica. Efectivamente se pudieron detectar las regiones importantes de la molécula como las regiones más conservadas, pero no resultó posible interpretar a qué se debía que fuesen importantes. Pero este fracaso se vió compensado con creces cuando los datos obtenidos mediante este enfoque comparado, se utilizaron para investigar la marcha seguida por la evolución de una misma proteína en diferentes especies. Se pudo así profundizar como nunca se había podido hacer, en los mecanismos a través de los cuales ha tenido lugar la evolución al nivel molecular. Además los datos convenientemente elaborados sirvieron para establecer árboles filogenéticos, que en el caso de los organismos superiores se pudo comprobar que coincidían muy bien con los obtenidos utilizando caracteres morfológicos. Con esta confianza, el mismo enfoque molecular se pudo aplicar a las bacterias, en la que los criterios morfológicos no habían sido capaces de elaborar árboles filogenéticos apropiados.

Por fin, los intentos de comprender la función a través de la estructura tuvieron éxito con las enzimas. No mucho tiempo después de que se estableciese las de mioglobina y hemoglobina, se determinó la estructura tridimensional de la primera enzima, la lisozina. Ya con esta enzima fue posible esclarecer numerosos detalles acerca del mecanismo de la reacción que cataliza. Y lo mismo ha ocurrido desde entonces con gran variedad de enzimas. Pero para lograrlo otra vez fue necesario utilizar el enfoque comparativo. Sólo que ahora era preciso comparar la estructura en el espacio de la enzima libre, con su estructura con ligandos apropiados unidos a sus centros específicos. De este modo, en los casos mejor estudiados se cuenta con un conjunto de estructuras, que a modo de viñetas, describen todo el mecanismo de la reacción catalizada con detalla atómico. Y lo más interesante de todo es que los resultados así obtenidos vinieron a confirmar los mecanismos propuestos a partir de datos de carácter indirecto, reunidos tras muchos años de labor paciente de los enzimólogos.

Otra tarea de gran interés de la enzimología ha sido el estudio de la regulación de la actividad enzimática. Los primero hallazgos tuvieron lugar cuando, en los años 50, se inició por fin el estudio de las vías anabólicas que hasta entonces habían resistido los intentos realizados para poder esclarecerlas. Entonces el progreso fue en cambio, muy rápido. La razón que justificó el rápido avance fue la utilización del análisis genético, al tiempo que los métodos clásicos de la bioquímica. Fue en esa época en que se hicieron las primeras observaciones decisivas acerca de los dispositivos reguladores y

cuando, a través del fenómeno de la inhibición por el producto final de la vía, se reconoció la existencia de las enzimas reguladoras. Fue también entonces cuando se perfilaron las primeras observaciones sobre el control de la actividad enzimática por fosforilación y desfosforilación, en el caso de la fosforilasa del glucógeno. Con el paso de los años la regulación alóstérica y la regulación por interconversión llegarían a ser reconocidas como dos mecanismos reguladores básicos y de amplia distribución en la naturaleza. Por su parte la regulación hormonal abordó el problema de la naturaleza de los efectores y en ese campo se establecieron interesantes conexiones entre la interconversión enzimática como dispositivo generador de la respuesta y la unión de la molécula activa con el receptor, como estímulo. Fue también a fines de los 50, cuando se reconoció la existencia de cAMP como segundo mensajero. También en esos años Monod desarrollaba la intensa labor que le condujo a comienzos de los 60 a proponer el modelo del operón para explicar la regulación de la síntesis de proteínas en las bacterias y a mediados de la misma década, el primer modelo alóstérico que comprendía como un elemento esencial los cambios de conformación en la proteína reguladora. (Cadenas, 1978).

## ACCIÓN DE LOS GENES

Algunas de las funciones asignadas al material genético habían podido ser explicadas con referencia a la estructura de la doble hélice del DNA y la complementariedad de las bases, en una y otra cadena de la molécula. No obstante, se necesitaba una descripción molecular de cómo los genes gobiernan las actividades de cada día en las células que los albergan. La historia de cómo se logró este objetivo constituye uno de los capítulos más brillantes de la Biología Molecular, escrito por los investigadores más extraordinarios con que ha contado esta ciencia.

La síntesis de proteínas está en último término gobernada por el DNA. El proceso es muy complejo. En él interviene gran número de componentes. El control por parte del DNA también se ejerce de una forma complicada. Por eso, la etapa inicial en la elucidación de la síntesis de proteínas bajo el control genético, estuvo jalona da por acontecimientos consistentes en el establecimiento de conexiones, que suponían la implicación de distintas sustancias en el proceso.

Existe una afeción artrítica cuyos pacientes presentan un color de vino en la orina. Es la alcaptonuria, una enfermedad hereditaria. En 1902 Sir Archibald Garrod dedujo de un estudio de los antecedentes familiares de los pacientes aquejados por este desarreglo, que se transmite por un gen recesivo. Se trata de una mutación que determina un defecto metabólico. La sustancia coloreada normalmente se transforma, pero en estos pacientes no ocurre así y se excreta de forma ostensible en la orina. Garrod reunió varios ejemplos de este tipo y sugirió para describirlos el término de "inborn error of metabolism". Así quedaba señalada por primera vez una conexión irrefutable entre genes y enzimas. Pero las ideas eran muy avanzadas para la época y el concepto de que los genes actúan directamente sobre el metabolismo no tuvo amplia aceptación hasta los años 40, gracias a los experimentos con mutantes bioquímicos de *Neurospora crassa* realizados por G. W. Beadle y E. L. Tatum.

Como resultado y resumen de la interpretación de estos trabajos se elaboró el aforismo un gen: una enzima. Con él se quería expresar que la función primaria de cada gen consiste en dirigir la síntesis de una enzima y de ese modo, controlar una reacción metabólica concreta.

Existe una anemia que prevalece en ciertas regiones de África. Se denomina anemia falciforme, porque los glóbulos rojos de los pacientes de esta grave enfermedad adoptan una forma de hoz muy característica. En 1949 Pauling junto con H. Itano y S. J. Singer probaron que esta anemia se debe a la posesión de una hemoglobina defectuosa que tiene una carga eléctrica menor que la normal, como se evidencia por medio de la electroforesis. Describieron el defecto como "molecular disease". Siete años más tarde V. M. Ingram conseguía probar que la molécula de la hemoglobina en la anemia falciforme difiere de la normal solamente en la sustitución de un residuo del aminoácido glutamato por el de valina. Nunca antes se había probado que una mutación consistiese en la sustitución de un residuo por otro en una cadena polipeptídica. Así se disponía de una pista que indicaba cómo un gen logra controlar la síntesis de una enzima. Siguiéndola se llegó finalmente a la conclusión de que lo que hace el gen es determinar la secuencia de la proteína, cuya síntesis gobierna. Pero el DNA no lo hace directamente, requiere el concurso de diferentes tipos de RNA, la otra clase de ácido nucleico. Fueron principalmente los trabajos de J. Brachet y de T. Caspersson los que a mediados de los 40 y como resultado de la aplicación de las técnicas histoquímicas, señalaron que las células implicadas activamente en la síntesis de proteínas tienen un alto contenido en RNA. Así los investigadores comenzaron a familiarizarse con la implicación del RNA en la síntesis de proteínas. Pero antes de que se pudiese llegar a un conocimiento detallado de en qué consiste ese papel, se necesitó disponer de sistemas *in vitro* para la síntesis de proteínas.

Zamecnik y su grupo obtuvieron en Boston en 1954 el primer sistema desprovisto de células capaz de incorporar *in vitro* aminoácidos en una cadena lineal, mediante la formación de enlaces peptídicos. El propio sistema reveló la participación de los RNAs de transferencia, la necesidad de GTP y que la síntesis tiene lugar en los ribosomas ricos en RNA. Pero no pudo mostrar la existencia del RNA mensajero, que lleva la información genética desde el DNA a los ribosomas. Fueron precisamente los experimentos de Jacob y Monod acerca de la regulación de la síntesis de proteínas en las bacterias los que pusieron de manifiesto la necesidad de la existencia de un mensajero y revelaron sus propiedades básicas. La evidencia era entonces indirecta pero poco después los experimentos de E. Volkin y L. Astrachan primero y los de Spiegelman después probarían en 1961 la existencia del RNA y mostrarían, a través de la comparación de la composición unos y a través de la hibridación otros, que muy probablemente el mensajero tenía una secuencia, que era transcripción de la secuencia del DNA, tal como se había postulado. El último paso, el de la acción biológica misma, se dio finalmente por M. Nirenberg y J. Matthaei que reseñaron en el Congreso Internacional de Bioquímica de 1961 en Moscú, que mensajeros naturales y artificiales estimulaban la síntesis de proteínas en un sistema *in vitro*. (Nirenberg y Matthaei, 1961).

El propio sistema descrito por Niremberg y Matthaei le sirvió a su grupo y al de S. Ochoa para sentar las bases de la correspondencia entre tripletes de bases en el mensajero y aminoácidos incorporados en la proteína. Esta correspondencia descrita como código genético, se obtuvo mediante ingeniosos procedimientos, que se adelantaron a la realización mediante la secuenciación de una proteína y del gen que la codifica, que hubiera sido el método más directo, pero entonces más difícil. El progreso fue rápido y pronto se reveló en todos sus pormenores el código, que fue ampliamente aclamado como depositario del secreto de la vida.

Desde que se desarrollaron los primeros sistemas *in vitro* para la síntesis de proteínas se produjo un progreso ininterrumpido en el conocimiento, no sólo de la implicación de más y más componentes en el proceso, sino de más detalles moleculares respecto del papel que juegan en el proceso en su conjunto en base a un conocimiento también exhaustivo de su estructura. El progreso afectó primero a los RNAs de transferencia que fueron las primeras moléculas de ácido nucleico secuenciadas. También fue una RNA de transferencia el primer RNA cuya estructura en el espacio se estableció por medio de difracción de rayos X. Más tarde el interés se trasladó a los ribosomas y sus componentes, las proteínas ribosómicas y los RNA ribosómicos. Se descubrieron factores proteínicos implicados en el proceso y se logró establecer el papel que juega el GTP. Por fin, y en un progreso aún no concluido, se está intentando tener una imagen lo más detallada posible de la organización de los componentes en el ribosoma, las interacciones de éste en las distintas etapas de la síntesis con los diferentes factores y los mecanismos moleculares de los que depende la catálisis, así como la fidelidad del proceso global de síntesis.

El esfuerzo realizado para conocer los componentes del aparato de síntesis de proteínas tuvo un fruto colateral de una enorme importancia, que habría de cambiar nada menos que nuestras ideas acerca del origen y evolución de todas las células en nuestro planeta. Para 1965 gracias a los esfuerzos de muchos investigadores y en particular de Zuckerkandl y Pauling ya se había comprendido que las macromoléculas encierran un registro de la historia evolutiva de la vida. Este resultaba especialmente pertinente para las bacterias cuya evolución no era posible desentrañar mediante la utilización de la morfología comparada. Woese y sus colaboradores iniciaron un proyecto para aclarar la filogenia de este grupo tan antiguo de organismos mediante la comparación de las macromoléculas de las diferentes especies. No abordaron sin embargo, la comparación de las secuencias de alguna proteína escogida. En 1972 comenzaron a estudiar RNAs ribosómicos convencidos de que debían ser mejores que las proteínas, debido a que parecen evolucionar más despacio que ellas y no alteran en nada el papel fisiológico que les fue asignado originalmente. Como entonces resultaba impráctico secuenciar hasta el menor de los RNA ribosómicos, optaron por elaborar catálogos del menor de los RNAs ribosómicos grandes. Para elaborar los catálogos sólo se fragmentaba el RNA y se secuenciaban los fragmentos, pero no se intentaba ordenar estos fragmentos en la secuencia total. El proyecto se llevó a cabo con una gran variedad de especies bacterianas y algunas eucarióticas. Pero los resultados totalmente inesperados forzaron a una revisión de los conceptos sobre la

evolución en sus estadios más tempranos, que aún no ha terminado de completarse, pero que ya ha trastocado todas las ideas que prevalecían con anterioridad.

En 1977 Woese y colaboradores propusieron la existencia de dos grupos distintos de bacterias, que no están emparentadas entre sí más estrechamente de lo que puedan estarlo con las células eucarióticas. Uno de los grupos lo denominaron Eubacterias y comprendía la mayor parte de las bacterias conocidas. El otro se designó como Arquebacterias, que originalmente sólo englobaba a las bacterias metanógenas. En 1978 se mostró que los organismos halófilos extremos de las salinas son también archebacterias y más tarde se probó lo mismo para un conjunto de organismos, que suelen vivir a altas temperaturas y tienen un metabolismo que depende del azufre, que comparten con las demás arquebacterias la posesión de lípidos muy poco usuales. El nombre de arquebacteria sugiere antigüedad. Sus propiedades parecen ser las que debieron tener los organismos que vivieron en la era arcaica y que prosperaban en condiciones que ahora nos parecen extremas. (Woese y Fox, 1977; Woese et al., 1978; Fox et al. 1980)

## INGENIERÍA DE LOS GENES

La Microbiología y sus métodos tuvieron gran trascendencia en el crecimiento de la Biología molecular. Ya hemos mencionado la escuela americana informacionista y a M. Delbrück, que formó el famoso grupo de los fagos, que protagonizó una época de progreso espectacular. También fueron bacterianos los sistemas *in vitro* que sirvieron para revelar la naturaleza del código genético. Fueron también estudios con bacterias los que llevaron a Jacob y Monod a su modelo del operón. En muchas de estas investigaciones las técnicas tuvieron originalmente carácter genético, para pasar después a ser sometidas al escrutinio de los métodos bioquímicos. Desde los años 60 el progreso ha sido continuado en nuestro conocimiento respecto de las manipulaciones a que se ve sometido en la célula el DNA. No sólo se ha logrado así un gran avance en los conocimientos, sino también un enorme progreso en las técnicas de manejo, así como en el arsenal de enzimas disponibles para realizarlo, hasta hacer del DNA la sustancia con la que se puede llevar a cabo el mayor número de alteraciones deliberadas. La conjunción de todas esas técnicas produjo en la década de los 70 el nacimiento de toda una nueva tecnología, que se ha denominado de diversas formas. Se habla de Ingeniería Genética, porque permite alterar a voluntad del acervo genético de los organismos, para nuestro provecho. Se llama tecnología del DNA recombinante, porque permite la preparación de moléculas de DNA mixtas, que a modo de quimeras, tienen porciones que provienen de especies distintas. Nos hemos acostumbrado a la clonación molecular, que permite la obtención de cualquier fragmento de DNA en que estemos interesados, en cualquier cantidad que estimemos necesaria. Esta tecnología permite cambiar todos los aspectos de nuestra vida en sus aplicaciones prácticas, pero de momento lo más importante que está haciendo es permitir el avance de la

Biología Molecular en todos los campos incluso en aquellas áreas que, como el desarrollo o el cerebro, se habían resistido más a sus ataques.

La clonación molecular permite obtener en cantidad cualquier gen tanto procariótico como eucariótico y después, el desarrollo de métodos avanzados por Gilbert y Maxam y por Sanger, hacen posible y relativamente sencilla la secuenciación del DNA. Es esta técnica la que revoluciona tantos campos de investigación. No sólo hace factible la secuenciación de una proteína, secuenciado el gen que la codifica. Permite hacer lo propio para proteínas totalmente desconocidas, que aún no se han aislado y que por ello, no se sabe qué función puedan desempeñar. Además facilita no sólo la secuencia de una proteína, sino la de su precursor, dando indicaciones acerca del camino que sigue su elaboración después de la síntesis. Permite conocer la secuencia y aún inferir la estructura de proteínas de membrana, como receptores, o canales. Hace posible acumular grandes cantidades de secuencias proteínicas que se pueden comparar, descubriendo analogías insospechadas, que indican conexiones evolutivas sumamente interesantes. En definitiva, ha cambiado la nueva tecnología los laboratorios, la naturaleza de las publicaciones científicas y promete una renovación total de la Biología Molecular, que, para profundizar en los mecanismos enzimáticos, puede aprovechar las técnicas de mutación dirigida que le permiten preparar proteínas con la estructura alterada; que permitan establecer la importancia de cualquier aspecto estructural, que nos propongamos ensayar, y quizás llegar así a la fabricación de superenzimas; que para comprender el funcionamiento del receptor para acetilcolina puede implantar en un óvulo los genes de sus distintos componentes o los genes manipulados y observar el comportamiento que resulta; que puede en fin, preparar vacunas, hormonas o proteínas del plasma en cantidad y a bajo precio, para su utilización en clínica.

Si hay algún campo en particular en que los cambios hayan sido más inesperados ha sido en el del conocimiento de los genes eucarióticos, que hasta el advenimiento de la donación molecular permanecía al abrigo de las técnicas de investigación. Una vez se comenzaron a aislar y estudiar algunos genes de virus eucarióticos primero y luego genes de células eucarióticas, se pudo constatar con enorme sorpresa que los genes eucarióticos no son continuos, como se creía y como son los de las bacterias, sino discontinuos, estando los tramos que codifican o exones, interrumpidos por tramos intercalados o intrones, que no lo hacen. La existencia de esos tramos no codificadores sirve para explicar, en parte, la existencia de un gran exceso de DNA en los organismos eucarióticos, por encima del que se precisa para codificar la información genética. Pero sobre todo cabe la posibilidad de que pueda explicar en gran medida la evolución acelerada de este tipo de organización celular. Es fácil que los intrones hayan simplificado la recombinación y también pueden haber permitido que la construcción de nuevas proteínas haya ocurrido barajando sus partes, hasta lograr la combinación mejor adaptada a las necesidades de la célula. De cualquier forma, resulta interesante preguntarse acerca de si los intrones ya existieron en los primeros organismos, siendo las eubacterias los únicos en renunciar a ellos después de que hubieron cumplido un brillante cometido, en aras de un ahorro

de recursos, mientras la célula antecesora de que provienen los eucariotas prefirió conservarlos. (Blake, 1983).

Parecía de sentido común, antes que se conociese la estructura génica de los eucariotas para compararla con la de los procariotas, suponer que estos últimos son más primitivos que los primeros, porque las células procarióticas son menos complejas y sus genomas más pequeños. Pero los descubrimientos de los últimos años han cambiado todo esto. Ya se sabía que la célula eucariótica ha tenido un origen mixto. Conforme a la teoría endosimbiótica, los orgánulos, como mitocondrias y cloroplasto, proceden de bacterias atrapadas por una célula ancestral, que en lugar de destruirlas, inició una convivencia de mutuo provecho. ¿Pero el núcleo y el citoplasma de qué clase de célula provienen? Sin duda debió tener organización procariótica por definición. ¿Pero de qué linaje proviene? Actualmente se vienen sucediendo las propuestas al respecto. La primera fue que eubacterias y arquebacterias se separaron a la vez que lo hizo la célula de la que se originarían núcleo y citoplasma. Después fue la de la que las células eucarióticas provienen de arquebacterias por su núcleo y citoplasma y de eubacterias por sus orgánulos. Incluso se ha propuesto que los organismos originales fuesen las arquebacterias, de cuya rama de metanógenos y halófilos extremos podían provenir las eubacterias, mientras núcleo y citoplasma podrían haberlo hecho de la rama de las arquebacterias dependientes del azufre. No conocemos aún la respuesta, pero sin duda está encerrada en las secuencias de DNA y su análisis nos revelará en último término de dónde venimos. ¿Pero, y lo que somos?

## SECUENCIACIÓN DEL DNA

Desde hace poco más de un año se vienen celebrando reuniones en Estados Unidos casi mensualmente, para discutir sobre el proyecto de llevar a cabo la secuenciación completa del genoma humano.

Una de las primeras ocasiones en que se trató seriamente la cuestión fue el verano del año pasado en la Gordon Conference sobre Genética Molecular. Casi al mismo tiempo Robert Sinsheimer de la Universidad de California en Santa Cruz convocaba una reunión informal sobre le mismo tema. La conclusión en ambos casos fue la misma esencialmente. El proyecto aunque gigantesco en su magnitud, era técnicamente factible.

Luego pasó a tomar la iniciativa el Departamento de Energía de Estado Unidos (DOE) que ha venido gastando dinero abundante en el estudio de los efectos de la energía sobre la genética humana (\$2.000 millones en los últimos años). Charles Lisi, el nuevo director de la Oficina de Investigación sobre salud y ambiente (OHER) de la DOE patrocinó en marzo un workshop en Santa Fe bajo la presidencia de Frank Ruddle de Yale University. En esa reunión la discusión ya no fue acerca de si puede secuenciar todo el genoma humano, sino de cómo hacerlo. Se consideró que el trabajo de organización es tan importante como la propia tecnología de secuenciación. Por eso se aconsejó hacer un esfuerzo, no sólo en el intento de lograr métodos de secuenciación más simples, más rápidos y menos costosos, sino también y muy especialmente, en el desarrollo de nuevas técnicas para el manejo de datos.

Para estos propósitos la OHER parece que podría invertir unos \$ 20 millones en los próximos tres años.

En ese workshop Walter Gilbert dijo con entusiasmo que: “*la secuencia completa del genoma humano constituye el Santo Grial de la Genética Humana*” (Lewin, 1986). Hace seis años, cuando se le concedió el premio Nobel ya afirmó que “*las secuencias del DNA son las estructuras definitivas de la Biología Molecular. No hay nada más primitivo. Las preguntas se formulan allí en último término*” (Kolata, 1980).

Walter Gilbert compartió con Paul Berg y Frederick Sanger el premio Nobel de Química en 1980. Conforme a la Real Academia Sueca el premio le fue concedido a Berg “*por sus estudios fundamentales sobre la bioquímica de los ácidos nucleicos, con particular referencia al DNA recombinante. Berg fue el primer investigador que construyó una molécula de DNA recombinante, es decir, una molécula que contiene partes de DNA de diferentes especies. Su pionero experimento ha conducido a una nueva tecnología que se suele llamar Ingeniería Genética*”. Sanger y Gilbert fueron galardonados por sus descubrimientos de nuevos métodos para secuenciar DNA. Estos métodos hicieron que la secuenciación del DNA pasase a ser más fácil y mucho más exacta que la secuenciación de proteínas.

En 1974 Winston A. Salser decía en una revisión sobre las técnicas de secuenciación de DNA (Salser, 1974) que si varios laboratorios uniesen sus esfuerzos quizás se pudiera secuenciar el genoma del SV40. Aunque ya apuntaba en esa revisión la revolución que se avecinaba con la utilización de la clonación molecular, señalaba como imposible la secuenciación del genoma humano. Recordando que una célula humana diploide contiene unos 5pg de DNA o sea aproximadamente un millón de veces más DNA que el fago SV40, el tamaño del genoma hace impráctico purificar una sola región para secuenciarla. Incluso si esa secuencia se pudiera purificar con un rendimiento del 100%, la cantidad de DNA de partida necesaria para proporcionar unos pocos microgramos de un fragmento de 1000 bp sería enorme. No obstante las cosas han cambiado mucho en la última década.

Sólo hace nueve años que se publicó en la revista NATURE la primera secuencia de un DNA humano (fue la del gen que codifica la somatotropina coriónica) y ahora no transcurre una semana sin que se describa alguna. Desde entonces en unos 2.500 artículos se han señalado unas 500 secuencias humanas diferentes, más muchos otros fragmentos. Se trata de un aumento exponencial en la producción científica sobre un tema concreto, como se ha dado en muchos otros casos con anterioridad. Algo parecido ocurrió, por ejemplo, cuando se descubrió la existencia y el papel fisiológico del cAMP a fines de la década de los 50. En todas las ocasiones el crecimiento remite después de transcurrido un tiempo. No puede ser de otro modo. Se ha calculado que si continuase el ritmo actual de crecimiento en la publicación de secuencias humanas para 1992 todas las publicaciones biológicas se ocuparían sólo de ese tema y para 1994 todas las publicaciones científicas harían lo propio.

Lamentablemente en la actual avalancha de publicaciones sobre secuencias humanas se da la redundancia en un grado excesivo. Se ha descrito 8 veces la clonación de la adenosina desaminasa y de las cadenas alfa y beta del receptor de las células T. Aunque en algunos casos se trate de formas alélicas, se ha publicado 17 veces la secuencia del oncogén c-myc y 33 veces la secuencia del gen para beta globina. Hoy publican estas secuencias unas 8 revistas científicas. Cabe esperar que de ahora en adelante esas secuencias vayan directamente a bancos de datos.

Carl Anderson del Brookhaven National Laboratory fue chairman de un workshop organizado en marzo de 1979 por la Rockefeller University. Allí se trató por primera vez la necesidad de crear un banco de datos para secuencias de DNA (DNA database). El proyecto se perfiló más en una reunión del National Institute for General Medical Sciences (NIGMS) de los National Institutes of Health (NIH) de los Estados Unidos, que tuvo lugar en julio de 1980. Pero el DNA database no comenzó a funcionar hasta 1982 cuando ya estaban en marcha proyectos similares en Europa (Nucleotide Sequence Data Library del European Molecular Biology Laboratory, EMBL, en Heidelberg) y en Japón. El contrato para DNA database no lo ganó la propuesta presentada por Margaret Dayhoff del Biomedical Research Institute de Washington D. C. que tiene una gran experiencia en el almacenamiento y manipulación de secuencias de proteínas. Lo obtuvo en cambio, la compañía Bolt, Beranek y Newman con sede en Cambridge, Mas., que tiene a su vez contrato con la compañía de Los Alamos National Laboratory, con Walter Goad como principal investigador. Los Alamos tenía interés desde hace mucho en el tema. Ya en 1960 celebró reuniones y conferencias sobre la cuestión de la comparación de secuencias. El contrato para el proyecto que se denomina GenBank fue por \$ 3 millones, con una duración de 3 años.

GenBank tenía en 1982 secuencias con 600.000 residuos, lo que suponía los 2/3 de los disponibles entonces. En 1985 contaba con secuencias con 3 millones de residuos. En 1986 este banco de datos, lo mismo que el europeo, han multiplicado por 25 los residuos almacenados en su primer año de trabajo. Con todo se han generado retrasos mucho mayores de los previstos. GenBank sólo tiene ahora el 19% de las secuencias obtenidas en 1985. No obstante se espera que para 1987 que es cuando vence el contrato, hayan doblado lo que hoy tienen almacenado. Esta inmensa tarea sólo se podrá lograr con medidas drásticas de simplificación en la elaboración de la información, así como mediante la introducción de prioridad para determinados temas, como ocurre con lo referente a las investigaciones sobre el SIDA.

En los bancos de datos una cuestión es el almacenamiento y oferta a los investigadores de las secuencias almacenadas y otra distinta la manipulación de la información que encierran.

Las secuencias del DNA contienen varios tipos de información. En primer lugar información sobre su propia estructura, no tanto como en el caso del RNA, pero mucha más de la que se creía. La estructura detallada de la doble hélice depende localmente de cuál sea la secuencia. En segundo término lleva información acerca de la estructura de otras macromoléculas: RNA y

proteínas. Además información sobre la manipulación por enzimas y otras proteínas de la estructura y la información contenida en el DNA. Por ultimo, también poseen mucha información acerca del camino seguido en la evolución de las distintas secuencias y aún sobre los mecanismos moleculares de los que ha dependido ese proceso. Estos distintos tipos de información están codificados conforme a claves diferentes, se debe entender que no existe sólo un código genético, término que se suele aplicar a la correspondencia que existe entre nucleótidos en el mensajero y aminoácidos en la proteína, cuya síntesis controla el mensajero, sino varios aplicables a los distintos tipos de señal para la manipulación de información. Estos códigos tienen un amplio campo de aplicabilidad, pero no son universales presentando variantes para distintos tipos de sistemas y especies.

También para manejar las secuencias de DNA se ha sentido la necesidad de crear instituciones centralizadas. En 1984 los NIH concedieron un contrato por 5 años y \$ 5'6 millones a una pequeña compañía de Palo Alto, Cal., IntelliGenetics, para la instalación de un Recurso Nacional de Ordenadores para Biología Molecular. La compañía surgió en 1980 de un proyecto de colaboración entre varios Departamentos de Stanford que se inició en 1975 con el nombre de MOLGEN. Se trataba de la aplicación de la metodología de la inteligencia artificial a la Biología Molecular. El Recurso que se denomina BIONET debe proporcionar acceso fácil a bancos de datos de secuencias de DNA y proteínas, una biblioteca de software sofisticado para la búsqueda, el contraste y la manipulación de secuencias y además y muy principalmente, debe desarrollar más software. Además uno de sus objetivos más importantes es el de contribuir a establecer una comunidad de biólogos moleculares, que se comunique constantemente y con rapidez. (Lewin, 1984).

Las primeras secuencias de DNA se publicaban en las revistas científicas, pero cuando alcanzaron un tamaño excesivo se planteó la necesidad de suprimir esa publicación, de modo que ahora van directamente al almacenamiento en el banco de datos, cumpliendo los investigadores determinadas formalidades en la presentación de la secuencia para, de este modo, ahorrar trabajo posterior.

La secuencia del DNA del fago IX174 DE 5.375 residuos ocupó 2 ½ páginas en la revista Nature (Sanger et al. , 1977). La secuencia del DNA del fago de T7 de 39.936 residuos ocupó 8 páginas de la revista Journal of Molecular Biology (Dunn y Studier, 1983) . La secuencia del DNA del genoma humano con 3 mil millones de residuos ocuparía 600.000 páginas, es decir unos 400 libros de 1.500 páginas. Es evidente que las secuencias largas han de ir directamente a los bancos de datos. Así ha ocurrido con la secuencia del DNA del virus de Epstein-Barr de 172.282 residuos que no se publicó (hubiera ocupado 34 páginas de una revista) sino que se depositó directamente en el EMBL database (Baer et al., 1984) .

## SECUENCIACIÓN DE GENOMAS

La propuesta de secuenciar en su totalidad el genoma humano no carece de precedentes. Ya se han logrado secuenciar otros genomas y de esa secuenciación se han obtenido datos muy importantes.

En un sentido estricto genoma es un conjunto de genes que poseen capacidad de replicación y expresión autónoma y que cuentan con una evolución propia. La autonomía supone autosuficiencia. Implica contar con todos los componentes necesarios para la síntesis de proteínas. El genoma debe contener genes estructurales, que codifiquen proteínas y los tres tipos de RNA, mensajero, ribosómico y de transferencia. Pero la autonomía no supone total independencia, porque la expresión de la información siempre está sujeta a influencias externas. El genoma además de genes debe contener múltiples señales, que interpretan y obedecen un conjunto de proteínas. La evolución opera sobre los genes y sobre las señales tratando de adaptar el organismo de la mejor manera posible, a las condiciones del medio en que desarrolla su actividad vital. Pero el término genoma se aplica también en un sentido amplio, para indicar un conjunto de genes cuya replicación es autónoma, pero cuya extensión no lo es y a los que no les falta su evolución propia. Así se habla de genomas de plásmidos, virus u orgánulos.

La organización celular eucariótica, que es la que poseen los organismos superiores, se caracteriza entre otras cosas, por poseer orgánulos, como mitocondrias y cloroplastos. Como vestigio de la vida independiente que un día disfrutaron, los orgánulos de las células eucarióticas cuentan con su propio DNA, que codifica una serie de componentes del mismo orgánulo y disponen de los medios necesarios para la replicación de su DNA y para la síntesis de proteínas codificadas en él.

En estos últimos años ha tenido lugar un considerable avance en la secuenciación de genomas de orgánulos. Se conoce la secuencia nüdeotídica completa del DNA mitocondrial de algunos organismos. La secuencia humana con 16.569 residuos fue la primera (Anderson et al. ,1981). Pero hoy se conocen las de res, ratón y rata, así como las de mosca del vinagre, rana, *Tripanosoma* y *Leishmania*, así como gran parte de la secuencia de erizo de mar, levadura, y hongos filamentosos.

Se ha podido comprobar que aunque las funciones genéticas esenciales del DNA mitocondrial se ha conservado en múltiples organismos de diferente nivel evolutivo, la estructura y organización de los genes y la forma de expresarse han evolucionado de diferente forma. Así mientras el genoma mitocondrial humano y en general, de mamífero, muestra una estructura muy compacta con los genes contiguos y carentes de intrones, los genes mitocondriales de levadura se hallan separados entre sí por regiones de DNA no informativo, algunas incluso fragmentadas por intrones. El DNA mitocondrial humano codifica dos RNAs ribosómicos, 22 tRNAs y 13 proteínas, todas ellas

integradas en sistemas de membrana del propio orgánulo. Parece que son las mismas proteínas las que están también codificadas en las mitocondrias de las demás especies.

La secuenciación del DNA mitocondrial y la secuenciación directa de proteínas codificadas en él permitieron establecer la clave genética utilizada en el orgánulo. Así surgió la sorpresa. El código genético que se estimaba que era universal, resultó no serlo. En las mitocondrias se utilizan variantes diferentes de ese código. Después por procedimientos análogos se ha podido establecer que existen también variantes en algunos protozoos y al menos en una bacteria.

La genética de las mitocondrias fue siempre casi imposible de analizar porque era muy difícil contactar con mutantes, porque siempre resultaban letales. El contar con la secuencia ha permitido entrar para las mitocondrias en la era de la genética por análisis del DNA. Ahora la mayor parte de los descubrimientos genéticos no se hacen con la genética tradicional, sino con técnicas bioquímicas. Hoy día el principal instrumento es la hibridación molecular de los ácidos nucleicos basada en la complementariedad de sus bases. Hoy los métodos de hibridación cuantitativa son, para los biólogos moleculares, lo que han sido los ensayos de actividad para los bioquímicos de proteínas. Esencialmente cualquier gen para el que exista una sonda de hibridación se puede clonar hasta obtenerlo en cantidades de miligramos, con las que es muy fácil su secuenciación. Además la genética por análisis del DNA se ha completado después de que se ha logrado sistemas de ensayo fiable, en lo que se expresa con exactitud cualquier gen aislado.

En la carrera hacia la secuenciación de genomas cada vez más complicados los logros más resonantes son los conseguidos muy recientemente por dos grupos de japoneses (Gray, 1986) . Ohyama et al. de la Universidad de Kyoto han secuenciado el genoma del cloroplasto de la hepática *Marchantia polymorpha*, una briofita. Se trata de un DNA con 121.024 residuos. Por su parte Shinozaki et al. , han secuenciado el DNA del cloroplasto de la planta de tabaco, una traqueofita. Ese DNA tiene 155.844. Las briofitas y las traqueofitas se separaron en la evolución hace unos 400 millones de años, pero ambos genomas tienen una organización muy similar. La diferencia de tamaño no afecta prácticamente a la información que contiene. Ambos genomas tienen unos 120 genes (genes para RNAs ribosómicos, genes para 30 ó 31 tRNAs genes para 55 proteínas identificadas y otros 30 genes para proteínas aún no identificadas) . Entre las proteínas identificadas hay 19 que son ribosómicas.

La organización del genoma de cloroplasto tiene claras semejanzas con los genomas bacterianos. Los genes con relación funcional, como en aquellos, aparecen arracimados. También la disposición de los genes para RNAs ribosómicos es la misma que en las bacterias. No obstante, los genes de los cloroplastos presentan intrones, como los genes eucarióticos.

## SECUENCIACIÓN DEL GENOMA HUMANO

La excitación generada por estos logros espectaculares en secuenciación de genomas de orgánulo es uno de los factores que más ha contribuido a la idea de abordar la secuenciación completa del genoma humano. Pero es conveniente reflexionar acerca de la magnitud del proyecto.

En la tabla I se muestra comparativamente datos acerca de los genomas de bacterias, hongos y mamíferos, que tienen tamaños crecientes. Como es fácil de comprender, cuando se preparan fragmentos del DNA de un genoma para clonarlos, el número de fragmentos es menor cuanto mayor sea el tamaño que tengan. Pero aún utilizando los fragmentos más grandes que hoy se pueden clonar, que son los de unas 40.000 bp, sería preciso clonar unos 75.000 fragmentos para poder secuenciar el genoma humano completo. La cifra es enorme pero probablemente sería preciso multiplicarla por un factor que puede estar entre 3 y 10 si se quiere trabajar con total garantía. Tomando el tamaño medio de los fragmentos como 50.000 bp la tabla II muestra los cromosomas de una mujer (aparecen los cromosomas sexuales X e Y) con la longitud de su DNA y el número de fragmentos que habría que clonar para poder secuenciarlo. Como se ve, el DNA más pequeño es el del cromosoma 21 que tiene 48 millones de pares bases y cuyo estudio requeriría clonar 2.900 fragmentos.

**TABLA I**  
NÚMERO DE FRAGMENTOS QUE ES PRECISO CLONAR

Tamaño de los fragmentos clonados (bp)	Genoma bacterias ( $2 \times 10^6$ bp)	Genoma hongos ( $2 \times 10^7$ bp)	Genoma mamíferos ( $3 \times 10^9$ bp)
$5 \times 10^3$	400	4.000	600.000
$10 \times 10^3$	200	2.000	300.000
$20 \times 10^3$	100	1.000	150.000
$40 \times 10^3$	50	500	75.000

Dahl et al., 1981.

**TABLA II**  
CONTENIDO EN DNA DE LOS CROMOSOMAS HUMANOS;  
LONGITUD DEL DNA DE CADA CROMOSOMA Y NÚMERO DE CLONES  
NECESARIOS PARA CADA CROMOSOMA

Cromosoma	Contenido en DNA (%) del contenido en los autosomas)	Longitud del DNA ( $bp \times 10^6$ )	Número de clones (50 kb por clon)
1	4.32	249	15.000
2	4.22	243	14.500
3	3.49	202	12.000

4	3.34	193	11.500
5	3.26	184	11.000
6	3.02	173	10.300
7	2.77	160	9.600
X	2.70	154	9.200
8	2.55	146	8.700
9	2.37	137	8.200
10	2.33	136	8.200
11	2.38	137	8.200
12	2.35	134	8.000
13	1.86	110	6.600
14	1.80	103	6.200
15	1.69	100	6.000
16	1.55	93	5.600
17	1.49	85	5.000
18	1.40	81	4.800
20	1.20	67	4.000
19	1.08	62	3.700
Y	0.92	53	3.200
22	0.86	52	3.100
23	0.82	48	2.900

Mendelsohn et al., 1973

Actualmente una persona puede secuenciar 100.000 nucleótidos al año al precio de \$ 1 por nucleótido. Según esto parece que el proyecto de secuenciar el genoma humano con la tecnología actual costaría más de \$ 2 mil millones y requeriría unos 30.000 años-hombre. Al ritmo actual de secuenciación de 2 millones de nucleótidos por año requeriría entre mil y mil quinientos años. El coste de manejo de datos viene siendo de 3 céntimos por nucleótido almacenado, para el volumen de 6 millones de bases, aunque este presupuesto se ha visto desbordado porque requiere personal más cualificado de lo que se anticipaba. A este coste, solo el manejo de datos del proyecto debería suponer unos \$ 100 millones en 10 años. Para comparar cifras de coste digamos que la factura de *superconducting colliders* es de \$ 3 mil millones y de la *space station* es de \$ 8 mil millones. Además David Smith del DOE opina que le proyecto sólo supondría 300 años-hombre, costaría \$ 300 millones y tardaría sólo 10-15 años.

La reunión de este verano en Cold Spring Harbor fue el Simposio titulado Molecular Biology of *Homo sapiens*. Allí hubo mucha discusión acerca de si se trata de una aventura acertada el acometer ahora la secuenciación del genoma humano. (Lewin, 1986) .

Todos los investigadores están de acuerdo en que secuenciar el genoma humano es esencial. Como ha dicho Sydney Brenner director del Medical Research Council Laboratory of Molecular Biology “*cualquier persona interesada en la investigación médica debería estar interesado en la secuencia del genoma humano. Los que no lo estén, deberían ser considerados como sujetos que necesitan tratamiento médico*”. Paul Berg señalaba que hay que

preguntarse si el proyecto vale la pena, no en términos de dólares, sino en relación con el impacto que habrá que producir en el conjunto de la ciencia biológica.

En marzo de este año Renato Dulbecco razonaba en un comentario (Dulbecco, 1986) que si queremos saber sobre el cáncer nos debemos centrar en el genoma celular. Se presentan ahora según él dos opciones o tratar de descubrir los genes importantes en cáncer por un procedimiento fragmentario, o secuenciar el genoma completo de una especie seleccionada. Para Dulbecco no cabe duda que el método de elección es secuenciar el genoma completo. Sólo así se podrá contar con sondas para todos los genes y se les podrá clasificar en relación con su expresión en los distintos tipos de células. Esta clasificación habría de constituir la pieza clave para el estudio de la progresión de la transformación maligna. Tampoco tiene duda respecto de la especie en que se debería centrar el esfuerzo. Se sabe ya que hay diferencias en el control en especies distintas, de modo que se debería abordar la secuenciación del genoma humano.

Pero Dulbecco piensa que el conocimiento del genoma humano no sólo sería útil en el estudio del cáncer. Debería ser básico para el progreso de la fisiología y la patología humana. Por ejemplo, el conocimiento de la regulación de la expresión de genes individuales en distintos tipos de célula debería beneficiar muchos campos de investigación, como el estudio de desarrollo y el de la organización del sistema nervioso. El conocimiento adquirido se proyectaría rápidamente en aplicaciones terapéuticas muy diversas. El esfuerzo técnico habría de ser intenso, pero una mejora de las técnicas actuales por un factor de 50, permitiría completar el trabajo principal en 5 años. También indicaba que el esfuerzo debía ser nacional y aún mejor internacional ya que: "la secuencia del DNA humano es la realidad de nuestra especie y todo lo que ocurre en el mundo depende de esas secuencias".

Pero la cuestión es si es acertado acometer el proyecto ahora. Sydney Brenner opina que es prematuro. Eric Lander del Whitehead Institute de Cambridge cree que para el año 2000 de una forma u otra contaremos con la secuencia. Pero el proyecto en megaescala para lograrlo cambiaría para siempre la estructura de la investigación biológica. David Bostein del MIT cree que el proyecto pondría en peligro a todos, pero muy especialmente a los investigadores jóvenes. En su opinión lo primero que se necesita es un mapa físico del genoma formado por fragmentos de más de 40 kb, que se solapen. David Baltimore opina igual y cree que el mapa costaría unos \$ 10 millones y se tardaría en conseguir 3-5 años.

Parece que en el Japón se está considerando seriamente el proyecto de secuenciación de genoma humano. De hecho existe el proyecto en el Instituto Riken de Tokio dirigido por Eiichi Soeda. Creen que secuenciar el genoma humano sería importante por las mejoras en biotecnología que derivarán de él. Las nuevas tecnologías han hecho la nueva Biología Molecular que ha accedido a áreas prohibidas con anterioridad a su desarrollo. Los nuevos desarrollos abrirían aún más puertas.

Akijoshi Wada Profesor de Biofísica de la Universidad de Tokio opina que la secuenciación del genoma humano por 100 expertos costaría \$ 100 millones y tardaría 250 años. Pero en una “fábrica de secuenciación” costaría \$ 600 y requeriría sólo 30 años. Wada en colaboración con la industria y con la Agencia Japonesa de Ciencia y Tecnología está desarrollando desde 1981 un sistema automático para la secuenciación del DNA.

Seiko Instruments and Electronics ha desarrollado ya un robot que hace la separación de fragmentos para los procedimientos de secuenciación de Sanger y de Maxam-Gilbert. Toyo Soda Manufacturing Co. ha diseñado un extractor específico para DNA, con una alta resolución de bases. Fuji Photo Film Co. ha creado un sistema para la reproducción en masa de films para la electroforesis en gel. Por fin, Seiko y Hitachi Software Engineering Co. están compitiendo para desarrollar máquinas para leer los genes y escribir la secuencia.

Actualmente la Universidad de Tokio y el Instituto Riken de Tokio están invirtiendo \$ 3-4 millones en un esfuerzo para automatizar la secuenciación del DNA. Cuando todos estos avances de estas instituciones y de la industria se reúnan, se cuenta con que será posible secuenciar 1 millón de bases al día. De modo que se podrá acometer dentro de 2 años la secuenciación del cromosoma más pequeño del genoma humano, el cromosoma 21 con 48 millones de bases para ser completada en 5 meses.

A pesar de todo ello Nobuyoshi Shimizu de la Universidad de Keio de Tokio afirma que no tiene el propósito de secuenciar todo el genoma humano. Para lograrlo se necesita una organización de la que, opina, el Japón carece.

El julio pasado hubo una reunión de los NIH en Bethesda patrocinada por el Howard Hughes Medical Institute. En ella el presidente del Instituto, Donald Fredrickson señaló que se puede secuenciar al genoma humano y se va a llevar a cabo el proyecto, pero la cuestión es cómo hacerlo, cuándo, por quién, así como quién será el que pague la factura. Casi nadie estaba a favor de hacer el esfuerzo ahora. La dificultad radica en el dinero que costaría y el temor de que se retirara la financiación a otros temas de investigación. Leroy Hood opinó que sería un error acometer el trabajo ahora, porque en pocos años sin duda se habrá multiplicado la actual velocidad de secuenciación, pasando a ser uno o dos órdenes de magnitud más alta que la actual. De hecho Hood ha publicado ya la inscripción de un método de secuenciación de DNA (Smith et al., 1986) . Va a aparecer un secuenciador automático que Hood ha contribuido a desarrollar y que estará pronto disponible comercialmente.

Donald Friedrickson dijo que sería una gran equivocación embarcarse en un esfuerzo en gran escala para la secuenciación del genoma humano con las técnicas de andar por casa que ahora tenemos. En cambio si se hacen ahora las inversiones necesarias para desarrollar la tecnología en 5 años estaremos en condiciones de hacer el trabajo con mucha mayor eficacia y mucho menor costo.

De modo que parece que disminuye el impulso para secuenciar el genoma humano y es sustituido por un proyecto de realización del mapa físico. Para ello se necesita fragmentar el genoma humano, separar fragmentos de unas 40 kb, preparar una genoteca en cósmidos, que son los únicos vectores capaces de llevar fragmento de ese tamaño, y luego ordenarlos en cada uno de los cromosomas. Se entiende, sin embargo, que después vendrían otras dos fases. La segunda consistente en la secuenciación de todo o parte del genoma humano y la tercera que implicaría la completa comprensión de la regulación de la expresión de todos los genes humanos.

## MAPAS GENÉTICOS

Gran parte del trabajo de los científicos consiste en la elaboración de mapas. A medida que los instrumentos de exploración se hacen más penetrantes, van revelando detalles cada vez más pequeños de la región sometida a escrutinio.

Desde aproximadamente 1913 se sabe que los genes están situados de forma monodimensional a lo largo de los cromosomas. Además se suponía ya entonces que todos los individuos de la misma especie deberían tener sus genes dispuestos en sus cromosomas de una forma muy similar. Pero era preciso establecer la posición de los genes en los cromosomas. Primero el orden en que están. Luego el lugar que realmente ocupan. Ambas informaciones se presentan por medio de un mapa genético. Los mapas genéticos se elaboran a partir de datos sobre las distancias relativas entre los genes, admitiendo, como lo hacen los geógrafos, que las distancias sean aditivas. Para hacer cartografía genética se necesita contar con marcadores, que se sitúan en el mapa como referencia.

Los mapas genéticos clásicos estaban basados en la relación directa que existe entre la frecuencia de recombinación de los genes y las distancias relativas en el cromosoma. Para preparar un mapa se necesitaban mutantes, cruces y caracterización de una abundante progenie. Esto ha funcionado muy bien con bacterias y fue llevado a sus límites con fagos. Benzer consiguió así revelar la estructura detallada del gen. A finales de los años 50 el análisis genético era aún, más potente que la microscopía electrónica (o que la secuenciación del DNA que aún no se había desarrollado). Pero con eucariotas, como el hombre, el método fallaba, porque esta especie tiene tiempos largos de generación y pequeño número de crías.

Se definió la genética como la ciencia de la variación. Hasta que llegó la era de la genética molecular no podía haber genética sin variación. Sólo se podía suponer la existencia de un gen a través de la existencia de formas alternativas de un carácter genético determinado. Para 1977 ya se conocían unos 1.200 loci génicos como sede de otros tantos genes humanos (100 en el cromosoma X y 1.100 en los autosomas). De ellos, alrededor de 900 se habían caracterizado gracias a la existencia de enfermedades congénitas, la mayoría

poco frecuentes, debidas a mutaciones en los correspondientes loci. (McKusick y Ruddle, 1977) . Pero la genética molecular proporcionó otros medios para inferir la existencia de genes específicos. Una vez conocida la existencia de un polipéptido se puede suponer la existencia del gen que lo codifica, aunque no se conozcan variantes.

En la asignación de los genes a cromosomas específicos, supuso un gran avance el desarrollo de la genética de las células somáticas. La realización de mapas con este tipo de genética se necesita fraccionar el genoma, de modo que se puedan estudiar aislados cromosomas o fragmentos cromosómicos. El modo de lograrlo consiste en utilizar células de ratón como portadoras del material genético humano. Los cromosomas humanos se transfieren a células de ratón por fusión celular, inducida por virus o sustancias químicas. Luego se seleccionan las células híbridas, matando las células parentales aprovechando deficiencias que se compensan en las células híbridas. Cada célula híbrida conserva unos cromosomas y pierde otros, siendo relativamente simple establecer correlaciones entre la expresión de determinados genes y la presencia de un cromosoma. Así se había localizado unos 120 genes en el genoma humano hace unos 4 años. (Tunnacliffe et al. , 1984).

Hay dos clases extremas de mapa genético. Un mapa fenotípico en que una característica, como una actividad enzimática o una enfermedad congénita, se localizan en un cromosoma concreto. Y un mapa molecular donde se expresa la organización y la secuencia del DNA que compone los genes. Las distancias que manejan uno y otro tipo de mapa difieren en unas 100 veces.

El hacer el mapa de secuencias de DNA de hasta varios miles de bases de largo es ahora un ejercicio de rutina, que combina tres técnicas básicas. Primero la clonación de fragmentos de DNA, generalmente en células bacterianas. Luego la electroforesis de los fragmentos y por fin, la hibridación molecular. La donación de secuencias de DNA sirve una variedad de propósitos en el contexto de la elaboración de mapas de DNA. Pero su principal función consiste en descomponer el genoma de gran tamaño, en fragmentos de una longitud que permita su análisis, así como la obtención en cantidad y en forma pura. La gran ventaja de basar la confección de mapa en las secuencias clonadas es que proporciona un medio de recuperar fragmentos definidos del genoma, en una forma que se puede utilizar para establecer qué papel biológico desempeña (Southern, 1982). La preparación de mapas de centros de restricción determina la posición de los centros que reconocen cada restrictasa a lo largo de la cadena del DNA. Luego esos centros se pueden utilizar como puntos de referencia para medir las distancias a lo largo de la cadena de DNA. La hibridación molecular de DNA puede servir para identificar secuencias específicas. La combinación de la hibridación molecular con la electroforesis en gel de fragmentos de restricción, permite localizar la posición de secuencias específicas en el mapa de restricción.

La brecha entre una y otra técnica se logra cubrir por medio de la utilización de los polimorfismos en la longitud de los fragmentos de restricción (restriction fragment length polymorphisms, RFLP) . Son fáciles de detectar con

las técnicas normales de laboratorio. Ya se conocen varios cientos de estos centros discretos de variación en el genoma humano.

## UTILIDAD DE LOS POLIMORFISMOS

Los mapas genéticos se han venido haciendo siguiendo la herencia de la variación en caracteres concretos. Ahora se puede mirar la variación de secuencia en segmentos concretos del DNA. Hace unos 10 años que se descubrió que las variaciones de secuencia entre individuos se podía detectar con restrictasas. Si se daba la circunstancia de que las variaciones afectaban a los centros que reconocen las enzimas, que de ese modo, aparecían o desaparecían, se alteraba el tamaño de los fragmentos que resultaban de su acción. Los polimorfismos en la longitud de los fragmentos obtenidos por la acción de las restrictasas se pueden detectar con sondas de DNA clonado, que se hibridiza con esos fragmentos sea cual sea su tamaño (que se ha determinado independientemente por electroforesis en gel). Si un RFLP se hereda con una pauta igual que una enfermedad genética es posible establecer dónde reside el gen de la enfermedad, unos pocos millones de bases arriba y abajo. Y todo mejora notablemente si se cuenta con varios marcadores. De hecho bastaría con un marcador polimórfico por cada 20 millones de bases para disponer de un mapa adecuado.

Jean Dausset está llevando adelante un experimento cooperativo poco corriente en el Centro para el Estudio del Polimorfismo Humano en el College de France en París. El objetivo es la elaboración de un mapa detallado de marcadores genéticos que comprenda todos los cromosomas del genoma humano. Un mapa así permitiría a los investigadores acertar con los genes defectuosos responsables de enfermedades hereditarias, incluso en ausencia de información acerca de los efectos bioquímicos de las mutaciones. Según Víctor Kusick de la Facultad de Medicina de la Universidad John Hopkins afectan a la especie humana unas 1.300 enfermedades hereditarias. Para la mayoría de ellas no se ha logrado establecer los efectos celulares, ni los genes implicados. Ejemplos bien conocidos son la fibrosis cística, la enfermedad de Huntington, la distrofia muscular Duchenne y la enfermedad granulomatosa crónica. Pero una vez disponga del mapa de marcadores genéticos, las familias afectadas por una enfermedad congénita concreta podrán ver sometidas a estudio, para ver si la enfermedad se hereda en conjunción con alguno de los marcadores. Cuando los dos caracteres se hereden juntos es probable que estén cerca, en el mismo cromosoma. Así el marcador serviría de referencia para localizar el gen causante del problema e incluso podría permitir su aislamiento y caracterización. En el peor de los casos, el marcador se podría aprovechar para diagnosticar la enfermedad y para identificar los portadores. El mismo tipo de aplicación podría tener para enfermedades que, como la diabetes o la esquizofrenia, tienen sin duda un componente hereditario.

El proyecto de Dausset pretende la obtención de suficientes marcadores. Para ello se han obtenido linajes celulares permanentes de todos los miembros de 40 familias grandes que comprenden varias generaciones identificadas por

investigadores de todo el mundo. Muchas las aportó el grupo de Raymond White en el Howard Hughes Institute de la Facultad de Medicina de la Universidad de Utah. Dausset ha proporcionado muestras a laboratorios de diversos países, que tienen que disponer de una sonda de DNA clonado para uno de los RFLPs, cuya herencia deberán seguir en las 40 familias. Dausset recibe los datos y los compara con ordenador con los datos referentes a otros marcadores obtenidos por otros grupos para establecer el grado de ligadura entre los marcadores. Hasta ahora se vienen manejando unos 200 marcadores, aunque sólo una parte tienen suficiente polimorfismo para ser útiles.

Mucho del progreso reciente en el diagnóstico de enfermedades hereditarias se basa en la utilización de los RFLPs. La ventaja del procedimiento estriba en que permite detectar la presencia de una lesión hereditaria aún cuando la lesión misma no se haya caracterizado (Newmark, 1984). Para poder aprovechar la información para el diagnóstico se necesita uno al menos y preferiblemente varios de ellos, que estén suficientemente cerca de la posición cromosómica de la lesión, para que así sea mínima la probabilidad de que unos y otros se hayan separado por recombinación génica en la formación de los gametos. Los polimorfismos no son resultado de la lesión, sino que dependen de la variación preexistente en la población. El que un polimorfismo concreto se asocie a una lesión en una familia dada sólo depende de que existiese en el alelo ancestral, en que se produjo.

A veces se sabe dónde está la lesión y en qué consiste, como en el caso del diagnóstico prenatal y la detención de portadores de fenilcetonuria clásica, en que se cuenta con el gen humano clonado para la enzima fenilalanina hidroxilasa (Woo et al., 1983) . Pero para otros sigue siendo un misterio la naturaleza de una lesión. Así ocurre con la enfermedad de Huntington, para la que Gusella et al. han descubierto un marcador polimórfico (Gusella et al, 1983) para la que el tramo en que ésta tiene tamaño suficiente para codificar hasta 25 proteínas. En ocasiones la localización de la lesión en el mapa genético conduce eventualmente a la identificación de su naturaleza. El caso más interesante es el de la enfermedad granulomatosa crónica (Royer-Pokora et al., 1986). Se ha conseguido clonar el gen que es anormal en la afección fagocítica en su forma ligada al cromosoma X. La clonación se ha hecho sin referencia a una proteína específica, basándose en cambio, en la posición en el mapa cromosómico. La secuencia nucleotídica del DNA clonado predice que debe codificar una proteína de al menos 468 residuos, sin homología con ninguna proteína conocida y cuya función por tanto, sigue siendo un misterio. No obstante, presenta entre los residuos 65 y 92 un tramo hidrofóbico que podría constituir un dominio transmembrana. De modo que probablemente se trate de proteína integral de membrana. Lo que sí está probado es que es defectuosa en cuatro pacientes de la enfermedad.

De cualquier forma el diagnóstico para personas que no se conoce si son o no portadores de la enfermedad, debe ser sólo llevado a cabo con el consentimiento de las mismas. Hace poco hubo una solicitud de realizar el ensayo para ver si portaba la enfermedad de Huntington una niña de 2 meses que iba a ser adoptada y fue rechazada sobre esa base. Hoy existen ya marcadores para un cierto número de afecciones hereditarias como distrofia

muscular Duchenne, fibrosis cística, enfermedad renal cíclica y se están buscando afanosamente marcadores para enfermedades como la de Alzheimer, la depresión y ciertos tipos de cáncer. Pero se comienza a realizar ensayos para establecer presencia de ciertas condiciones a la hora de considerar un candidato para un puesto en la industria. La aplicación de este *genetic screening* en opinión de muchos va a constituir un grave atentado a la intimidad y a la igualdad de oportunidades en el empleo. Existe ya un test por ejemplo para detectar deficiencias en alfa-uno-antitripsina para trabajadores que estén en contacto con asbestos o polvo de algodón.

Los polimorfismos del DNA han revolucionado el análisis genético humano pero los RFLPs no se detectan con frecuencia y su grado de polimorfismo es bajo. Está claro que el análisis genético se simplificaría mucho, si se pudiera contar con sondas para regiones hipervariables del genoma humano, con variación multialélica y altas heterozigosidades. La primera región así la aislaron por casualidad Wyman y White hace seis años (Wyman y White, 1980) . Luego también por azar se han hallado otras regiones hipervariables asociadas al gen para la insulina humana, genes para alfa globina y el oncogen c-Ha-ras-1. En todos los casos la región hipervariable consiste en repeticiones en tandem de una corta secuencia (minisatélite) . El polimorfismo resulta de diferencias alélicas en el número de repeticiones, que se deben originar por intercambios desiguales en los procesos de mitosis o meiosis o por corrimiento del DNA durante su replicación. La variación resultante en la longitud de minisatélite se puede detectar utilizando cualquier restrictasa, que no corte la unidad repetida. Así se logran marcadores que se heredan de forma estable.

Los fragmentos obtenidos de los minisatélites hipervariables son similares a los debidos a RFLPs, pero sus longitudes varían, no porque se hayan alterado los centros de reconocimiento de restrictasas por sustitución de un nudétido por otro, sino debido a la variación en el número de secuencias repetidas dentro de los fragmentos. En comparación con los RFLPs convencionales que representan marcadores simples, estos nuevos polimorfismos tienen localizaciones múltiples en las que pueden servir de marcador simultáneamente. Por fin resultan superiores a los RFLPs convencionales en que son mucho más variables de modo que muchas veces la heterozigosidad de la población se acerca al 100% (Jeffrey et al. , 1985). Por añadidura la técnica de Jeffreys et al. tiene interesantes aplicaciones en Medicina forense (Gill et al., 1985) .

## ALGUNAS CONSIDERACIONES FINALES

En 1963 Sydney Brenner propuso al Medical Research Council desde su Laboratorio de Biología Molecular, un proyecto para estudiar la biología molecular del desarrollo de organismo pluricelular más sencillo posible: el gusano de 1 mm. de longitud y con una vida de 3 días y medio *Caenorhabditis elegans*. Como decía en el proyecto: “*nos gustaría atacar el problema del desarrollo celular... eligiendo el organismo diferenciado más sencillo posible y sometiéndolo a los métodos analíticos de la genética microbiana*”. El organismo consta de 959 células somáticas, de las cuales 302 forman el sistema nervioso. Cuando se propuso el trabajo muchos investigadores pensaron que estaba

adelantado para la época. Watson llegó a decir que en 20 años. Pero el proyecto se completó en 1983 (Lewin, 1984).

Ahora se conoce el camino seguido en el desarrollo por cada una de las células de *C. elegans*, incluidas las conexiones de las neuronas en su sistema nervioso. Y no parece tener ningún plan sencillo. Ni siquiera la existencia en el cuerpo del animal de una simetría bilateral se refleja en el desarrollo, que parece ser en cambio oportunista. Se han logrado un buen número de mutantes que afectan al desarrollo, con los que se espera poder profundizar en el conocimiento de su control.

Al concluir el proyecto Brenner comentaba que al comienzo se dijo que la respuesta para la comprensión del desarrollo iba a provenir de un conocimiento de los mecanismos moleculares del control genético. Pero los mecanismos moleculares son monótonos de tan simples y no nos van a decir lo que queremos saber. Tenemos que intentar descubrir los principios de organización. Como sucede que gran cantidad de componentes se montan en un mismo sitio, no creo que esos principios de organización estén incorporados en un dispositivo sencillo, como el código genético. Al principio se creyó que el desarrollo estaría impreso en el genoma como un programa parecido a la codificación de la estructura de las proteínas a través de su secuencia. Pero según Brenner, tendemos a hablar de forma laxa sobre programas genéticos, pero deberíamos ser cuidadosos con las implicaciones de esa expresión, incluso cuando se use metafóricamente. Ahora prefiere utilizar términos como representación o descripción internas y dice que *“la explicación total de todo organismo reside en él y uno siente que tiene que haber una gramática en alguna parte de su interior. En último término el organismo debe ser explicable en término de sus genes, simplemente porque la evolución ha ocurrido a través de alteraciones en el DNA. Pero la representación no será explícita. Se necesita entender la gramática del desarrollo para poder encontrarle su sentido”*.

Brenner trata de acalarar su punto de vista utilizando una analogía. La cabeza icosaédrica de un fago es un objeto geométrico preciso. Esa geometría se hereda. Así que debe haber una definición del icosaedro en algún lugar de su genoma. Pero no se encuentra ningún gen que diga: hacer un icosaedro. Es diferente de escribir un programa de ordenador para que dibuje un icosaedro. Para entender lo que significa especificar un icosaedro hay que entender primero los principios del ensamblaje molecular, la manera en que las proteínas del capsidio interactúan y se autoensamblan. El icosaedro está codificado de tal forma que la información está distribuida por todo el genoma completo. Es una buena imagen para entender que es lo que se necesita para hacer una mano o un hígado. Las especificaciones de estas estructuras están dispersas por todo el genoma. No se trata de una información clara y secuencial como la de juntar aminoácidos para hacer una proteína.

Todo esto es muy pertinente con el tema de la secuenciación del genoma humano completo y qué se podrá hacer con la información que nos proporcione. No es que la posesión de la secuencia en sí misma vaya a decir nada. Pero deberá facilitar el camino hacia la comprensión definitiva del

desarrollo y las funciones del organismo humano. La secuenciación se realizará sin duda y deberá de catalizar un nuevo renacimiento científico centrado en la especie humana que pasará a ser así la mejor conocida y la que sirva de referencia para el conocimiento de las demás especies animales. Proyectos como el de la secuenciación completa del genoma humano pueden ser discutidos en términos de los costos que van a representar, como ha ocurrido con la investigación espacial, pero es cuestión de tiempo y siempre más breve de lo anticipado. En el caso de la biología molecular todos coinciden en que encierra la posibilidad de la comprensión de los sistemas biológicos a su nivel básico que es el molecular. Como ha dicho Brenner: “*la biología molecular es el arte de lo inevitable, si trabajas en ella es inevitable que encuentres como funciona, por lo menos al final*”.