

# PROTEÓMICA: ¿QUÉ SON Y PARA QUÉ SIRVEN LAS PROTEÍNAS?

ÁNGEL MARTÍN MUNICIO  
Real Academia de Ciencias

## INTRODUCCIÓN

La historia de la ciencia nos muestra el gran periodo de maduración del concepto de *proteína*, que se extiende desde 1838 hasta el advenimiento de la *bioquímica* como campo independiente del conocimiento, al finalizar el primer cuarto del siglo XX. A partir de entonces las proteínas fueron consideradas entidades moleculares fijas dotadas de propiedades químicas y físicas singulares, basadas en la existencia de tres niveles estructurales –estructura primaria, secundaria y terciaria–. A su definición contribuyeron sobremanera las nuevas ideas y métodos que aportó el nacimiento de la *biología molecular*. El establecimiento del concepto de las relaciones *estructura-función* ha sido extraordinariamente fructífero para los estudios de la *actividad* de las proteínas, hasta que, con el conocimiento de los genomas de una colección de especies, el hombre incluido, se han elaborado nuevas ideas y nuevos procedimientos para llegar a la comprensión de la fun-

ción de los genes. Y es precisamente este nuevo campo de conocimiento el que ha comenzado a conocerse como *proteómica*. La relación histórica de estas etapas se recoge en la tabla I.

Ha sido así, pues, como, en el seno de la biología molecular, gracias al amplio conjunto de conocimientos sobre la *síntesis, estructura, propiedades, función y aplicaciones* de las *proteínas* desarrollado en la segunda mitad del siglo XX, y tras la comprensión de la *función de los genes* y de los abundantes datos sobre las *secuencias genómicas* de distintas especies, el hombre incluido, y de manera colindante o, quizás, solapándose, se ha desarrollado recientemente un nuevo ámbito de conocimiento: el complejo estudio del *análisis en gran escala de las proteínas* y de sus *relaciones funcionales*. Este estudio se conoce actualmente, como se acaba de subrayar, con el nombre de *proteómica*.

A pesar de la acumulación en las bases de datos de enormes cantidades de información acerca de las secuencias de DNA, ello no era suficiente para la elucidación de la función biológica. Y, por otro lado, no existe una estricta relación lineal entre los *genes* y el complemento de *proteínas* –el *proteoma*– de una célula.

Efectivamente, el conocimiento actual de las secuencias genómicas de diversas especies permite enfocar la in-

Tabla I

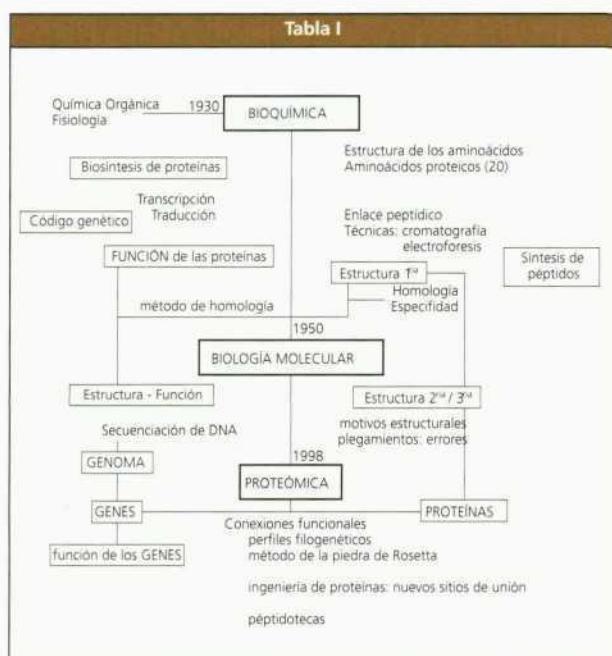


Tabla II

1. La biología está pasando de ser una ciencia con fuerte componente analítico a ser una ciencia cuyo énfasis se desplaza hacia la integración de informaciones globales sobre la función celular a niveles diversos, para producir conocimientos relevantes sobre los procesos vitales –obtener información sobre el entrampado global de las interacciones celulares.
2. Las eras genómica y proteómica se caracterizan por los procesos de automatización y robotización (secuenciación, RMN, rayos X), micromatrices de DNA (secuenciación por hibridación, chips de DNA), interacciones entre proteínas mediante técnicas experimentales y métodos computacionales (bioinformática).
3. Las matrices de DNA se basan en la idea de disponer en una pantalla de hibridación (chips de DNA) los 63 536 octámeros posibles en forma de una matriz reticular. Y un fragmento de DNA de varios centenares de bases hibridará solamente algunos de los octámeros, con lo que la secuencia podrá deducirse de forma computacional.

vestigación hacia la *comprensión de la función de los genes*; en particular, de los 40 000 o 50 000 genes que, aproximadamente, posee el hombre, y cuyos primeros borradores de las secuencias prácticamente completas —obtenidos por la empresa privada Celera y el sistema público americano— se acaban de hacer públicos a comienzos de 2001, aunque se desconozca la función de alrededor de sus dos terceras partes. Función compleja que, de otro lado, constituye la base entera del comportamiento celular.

Con estas ideas generales se pueden ya anticipar resumidas las diferencias fundamentales, que aparecen en la tabla II, entre los ya tradicionales —aunque recientes— estudios de las proteínas y los actuales objetivos que cumple la *proteómica*. Estas diferencias van a contemplarse, en primer término, desde un punto de vista histórico y, después, desde el ángulo de la ciencia biológica actual, al lado de las consecuencias que pueden deducirse en los aspectos curriculares y de la política de la ciencia.

El conjunto de temas que potencialmente componen los tratamientos de la *proteómica* aparece en la tabla III.

Tabla III
❑ Análisis estructurales de las proteínas: estructuras primaria, secundaria y terciaria
❑ Identificación de proteínas:
❑ Función de proteínas:
❑ Modificaciones postraduccionales: fosforilación, glicosilación, etc.
❑ Separación diferencial bidimensional en geles:
❑ Interacciones proteína-proteína:
❑ Medicina molecular:

## ANTECEDENTES

Era, sin embargo, a comienzos del siglo XX, en 1903, cuando Rodríguez Carracido, en su *Tratado de Química Biológica*, en el capítulo dedicado a los llamados «albuminoides», hacía los siguientes comentarios:

Grandes analogías presentan las numerosas sustancias incluidas en este grupo, y, sin embargo, su definición es tan vaga, que tiene todos los caracteres de agrupación convencional, de la que nada se puede afirmar sin advertir excepciones (...). Los albuminoides son los compuestos que forman la sustancia esencial y fundamental de la organización, por lo que se denominan también materias proteicas, como por su semejanza a la albúmina de la clara de huevo se llaman materias albuminoideas. Son compuestos generalmente sulfonitrogenados cuya composición centesimal presenta considerables variaciones. Su carácter de coloides cada vez tiene menor valor para caracterizarlos, porque dista mucho de serles exclusivo y, por otra parte, parece ser accidental, como en los demás cuerpos en los que se presenta, ante las proporciones en que se va revelando como posible la cristalización de los albuminoides. Respecto a la función química, Berthelot los considera amidas complejas y Gautier, nitrilos, también complejos, pero estas opiniones son por todo extremo insuficientes para definirlos (...). La extraordinaria flexibilidad de la materia constitutiva de los organismos para responder a todas las excitaciones del medio ambiente, reside, no sólo en lo inestable de la molécula resultado de su enorme magnitud, sino también en la coexistencia de numerosos grupos funcionales, que siempre han de presentar un punto de ataque a cualquiera de las acciones físico-químicas que sobre ellas incidan. Como el proceso biológico es resultante de acciones mecánicas, físicas y químicas, la doble escala progresiva y regresiva de los albuminoides es la síntesis de toda la Química.

Y al describir el apartado correspondiente a la constitución de los albuminoides, afirma:

Las diferencias correspondientes a las varias reacciones coloridas indican lo numerosos y variados que son los grupos moleculares integrantes de las multiformes materias proteicas, y explican las dificultades que presenta el discernir los eslabones de su enmarañada cadena, la cual, además, es tan delicada que los reactivos químicos, aun los de menor energía, no obran como escalpelos que disecan sus moléculas, sino como hachas que las destrozan.

Y acerca de su magnitud molecular dice:

Para la resolución de este problema se han ensayado los procedimientos físicos, y principalmente el crioscópico; pero, los resultados presentan tantos motivos de incertidumbre, que, no obstante su indeterminación, sólo se toman en cuenta las magnitudes calculadas por los datos químicos. Lieberkühn, refiriendo a un átomo la proporción centesimal de azufre y corroborando este dato con el del análisis de un albuminato potásico, estableció el número 1 612 como peso molecular de la albúmina, representándola por la fórmula  $C_{72}H_{112}N_{18}O_{2}S$ , la cual indudablemente es muy baja, pero todavía se usa en las reacciones bioquímicas cuando se expresan metamorfosis de la albúmina, en las que, por no ser necesario, o por ignorancia del proceso, sólo se simbolizan las relaciones cuantitativas que manifiestan la persistencia del peso de la materia transformada.

Apartado que concluye con el comentario:

De todo lo expuesto se infiere que las moléculas albu-minoideas son desmesuradas; pero que la determinación precisa de sus magnitudes dista mucho de verse realizada, y, por consiguiente, que falta el primer dato para el estudio sistemático de sus transformaciones, que es el conocimiento de las especies químicas.

Y en el capítulo referente a la formación de los albu-minoideos, Carracido afirma:

A las plantas que tienen clorofila les bastan sustancias minerales para su alimentación, y con ellas fabrican todos sus principios inmediatos, incluso los complicadísimos albu-minoideos. ¿Cuál es el mecanismo de esta asombrosa síntesis? La observación minuciosa de las condiciones en que la vida los elabora es la que puede dar la respuesta, pero antes de exponer los resultados obtenidos por este medio conviene anticipar para el mejor conocimiento de la índole del problema las tentativas de la experimentación en su afán de reproducir la obra de la Naturaleza.

Estos pocos retazos de la literatura científica del primer cuarto del siglo XX evidencian la confusa situación que el concepto de *proteína* ofrecía aún en esta época. Tras el descubrimiento por el holandés Mulder, en 1838, de las sustancias *primeras* entre los materiales de la vida –las *proteínas*– en numerosas fuentes animales y vegetales, esta época dejaba atrás todo un siglo de doctrinas vitalistas y especulaciones pseudocientíficas basadas en la llamada biocoloidología. A este descubrimiento se refirió Liebig, una de las autoridades químicas de la época, al asegurar que «abrió un universo de nuevos descubrimientos»; pero ello no fue inconveniente para que, a la vez, manifestara su discrepancia con la idea de *proteína* como «sustancia molecular ordinaria cuya fórmula pudiera determinarse por los métodos ordinarios de la química». Y esta época se enfrentaba, además, con el establecimiento por Fischer de la *teoría peptídica*, que había de constituirse en el paradigma de todos los estudios conducentes al reconocimiento de las proteínas como *macromoléculas bien definidas*.

#### EN BUSCA DEL ARGUMENTO BIOLÓGICO UNIFICADOR

Durante casi un siglo, de 1820 a 1918, se fueron aislando e identificando sucesivamente, con diferentes técnicas de hidrólisis, cada uno de los aminoácidos constituyentes de las *proteínas*, de forma que la hipótesis química de la *teoría peptídica* quedaba del todo fortalecida, en busca del modo de asociación de los aminoácidos en la estructura global. Y, una vez afirmado el concepto de *macromolécula*, debido principalmente a las ideas y los métodos de Staudinger y Svedberg, en el segundo cuarto del siglo XX, faltaba un *argumento profundamente biológico* capaz de engrasar los *principios estructurales*, la *especificidad funcional* y los *esquemas de biosíntesis*.

Esta falta de *argumento biológico unificador* sirve para comprender que, a pesar del esfuerzo de los más ilustres químicos orgánicos de la época –Abderhalden, Willstätter, Karrer, etc.–, Osborne pudiera afirmar «como, tras una abrumadora cantidad de trabajo, no se haya prácticamente progresado en los aspectos fundamentales de la estructura de la molécula de las proteínas»; argumentación biológica que, de otro lado, se vio soportada, en la década de los cuarenta del siglo XX, por el desarrollo de nuevos conceptos y métodos para su investigación, basados fundamentalmente en las técnicas isotópicas y en los variados procedimientos cromatográficos.

Para situarnos correctamente en el tiempo de estos hechos, recordemos que eran ya los comienzos de la segunda mitad del siglo cuando Pauling publicaba una colección de datos acerca de las *distancias interatómicas*, los *ángulos de enlace*, la *coplanaridad del sistema amídico* –CONH– y otras propiedades de los polipéptidos sintéticos, como ideas precursoras de la *estabilización de la estructura de las macromoléculas por los enlaces de hidrógeno* y, a la vez, de las *configuraciones helicoidales* de las cadenas. Hubo que esperar, sin embargo, a la filigrana experimental químico-enzimática de los trabajos de Sanger sobre la estructura de la insulina, de 1952, para probar definitivamente la *teoría peptídica*. En aquellos años, de todas maneras, la biosíntesis del enlace peptídico se buscaba a través de la reversibilidad de su escisión hidrolítica enzimática y del desplazamiento del equilibrio por variación de las propiedades físicas de los productos sinterizados. Y, aunque cada vez más cercano, seguía faltando el *argumento biológico unificador de ideas*.

Seguramente, la etapa definitiva en esta búsqueda surgió del conocimiento de la estructura y función de los *ácidos nucleicos*; a propósito de lo cual cabe señalar que, si la previa experiencia química y metodológica sobre la estructura de las proteínas pudo ser de cierta aplicación al conocimiento de los ácidos nucleicos, tuvieron que ser su *nativa heterogeneidad de composición, tamaño y secuencia de bases* y su relación con la síntesis de las cadenas polipeptídicas los conceptos moleculares inicialmente soportadores de la *añorada idea biológica*. Idea que conectaba la *información contenida en el material genético y la especificidad de la secuencia de las proteínas*.

#### LA NUEVA CONCEPCIÓN BIOQUÍMICA

A la maduración de la idea anterior contribuyó una colección de aportaciones tales como la síntesis orgánica de esteres fosfóricos, la idea de *derivados fosfóricos* con elevadas energías libres de hidrólisis, la *activación química* de los grupos carboxilo, y, en consecuencia, de los ácidos orgánicos y los aminoácidos, la averiguación de la estructura de la *coenzima A* y la introducción del concepto de *activación biológica*, la extensión de las técnicas de *difracción de rayos X*, y la correlación entre la cantidad de RNA y la velocidad de síntesis de proteínas. Aportaciones desde el

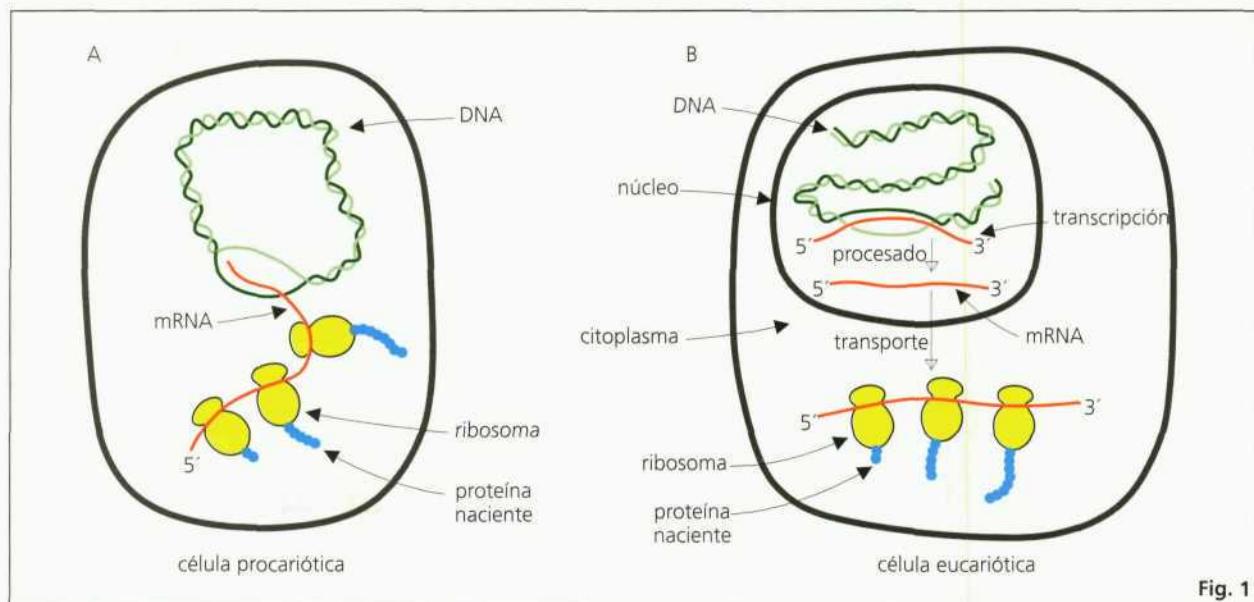


Fig. 1

lado de la química que se completaron con la utilización de tejidos, células y preparaciones del fraccionamiento celular para investigar el metabolismo intermedio y la incorporación de aminoácidos con las nuevas técnicas isotópicas.

Y, a todo esto, nos encontramos en la década de los cincuenta, posiblemente la más fructífera en la aparición del fantástico campo independiente del conocimiento que supuso la *bioquímica*. A partir de este momento, toda esta magnífica integración de conceptos y de métodos sirve ya a las propias circunstancias bioquímicas, con las tendencias o soportes que siempre tenderán hacia lo biológico o hacia lo químico. Del lado de lo biológico, las proteínas se conectan con el descubrimiento de los diferentes RNAs, las implicaciones genéticas del DNA y la naturaleza del código genético. Del lado de lo químico va a tener que ver con la interpretación de los numerosos mecanismos de isomerización, condensación y transposición; la aromatización, desaromatización y rotura de los anillos aromáticos; la apertura de ciclos; las reacciones de activación de fragmentos de 1C, 2C y 5C; la degradación C a C de las complicadas estructuras de esteroides, porfirinas y corinas; las reacciones de polimerización a melanina o caucho; y, a no dudarlo, la síntesis química de oligonucleótidos, tan decisiva en la averiguación del código genético. La figura 1 muestra, muy simplificadamente, los principales aspectos comparados de las etapas principales de la biosíntesis de las proteínas –transcripción y traducción– en las células procarióticas y las eucarióticas. La figura 2 resume los mecanismos del alargamiento sucesivo de la cadena polipeptídica en el seno de los ribosomas en los que se localizan las posiciones P y A, ajustadas por los correspondientes tripletes del mRNA. La síntesis se inicia por la colocación del f-metionil-tRNA<sub>f</sub> en el sitio P y la entrada en A del aminoacil-tRNA<sub>A</sub>. La acción de la peptidiltransferasa cataliza la formación del primer péptido en el sitio

A, ocupado por el aminoacil-tRNA<sub>A</sub>. A continuación, el peptidil-tRNA<sub>A</sub> recién formado se trasloca al sitio P, que se ha descargado del tRNA<sub>f</sub>, con lo que el sitio A queda libre para la entrada del segundo aminoácido bajo la forma del aminoacil-tRNA<sub>A</sub>, de forma que se puede comenzar un nuevo ciclo de alargamiento de la cadena peptídica.

#### LOS NIVELES ESTRUCTURALES DE LAS PROTEÍNAS. LA ESTRUCTURA PRIMARIA

La fácil aceptación por la bioquímica de los métodos y conceptos de la química y de la física hizo posible, en la década de los cincuenta, tanto la filigrana experimental de los trabajos de Sanger sobre la insulina, que probó definitivamente la teoría peptídica, como la medida de sus longitudes interatómicas y la de los ángulos de enlace, que, merced sobre todo a los trabajos iniciales de Pauling, demostraron la naturaleza planar de este grupo funcional peptídico (figura 3). Por primera vez, con Sanger, se definió la estructura de una proteína, la *insulina* (figura 4), pero, con todo lo que este hecho aportaba en aquel momento, fue mucho más relevante la introducción de una metodología original para la determinación de la *estructura primaria* de las cadenas polipeptídicas, conjugando los métodos químicos con los enzimáticos, la determinación de los puentes disulfuro, las hidrólisis parciales, las modificaciones específicas de los aminoácidos y la separación e identificación de numerosos productos de degradación. Todo este conjunto de facetas parciales de la química de proteínas no puede desvincularse de otra serie de hechos que contribuyeron decisivamente al impulso de la química moderna; se trata de las múltiples variantes de los procedimientos cromatográficos y electroforéticos de separación, con los que Tiselius, Martin y Synge, y Moore y Stein,

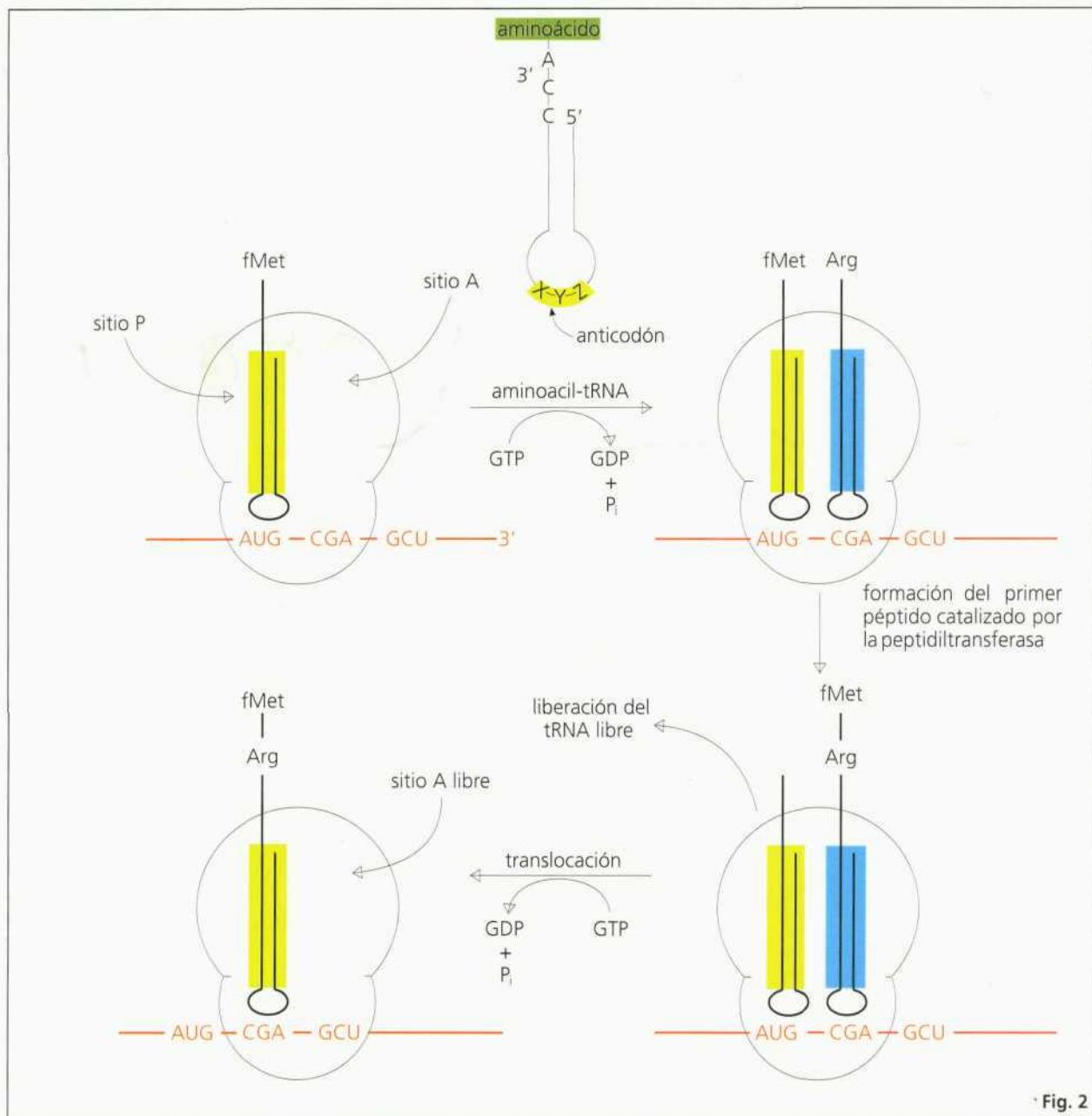
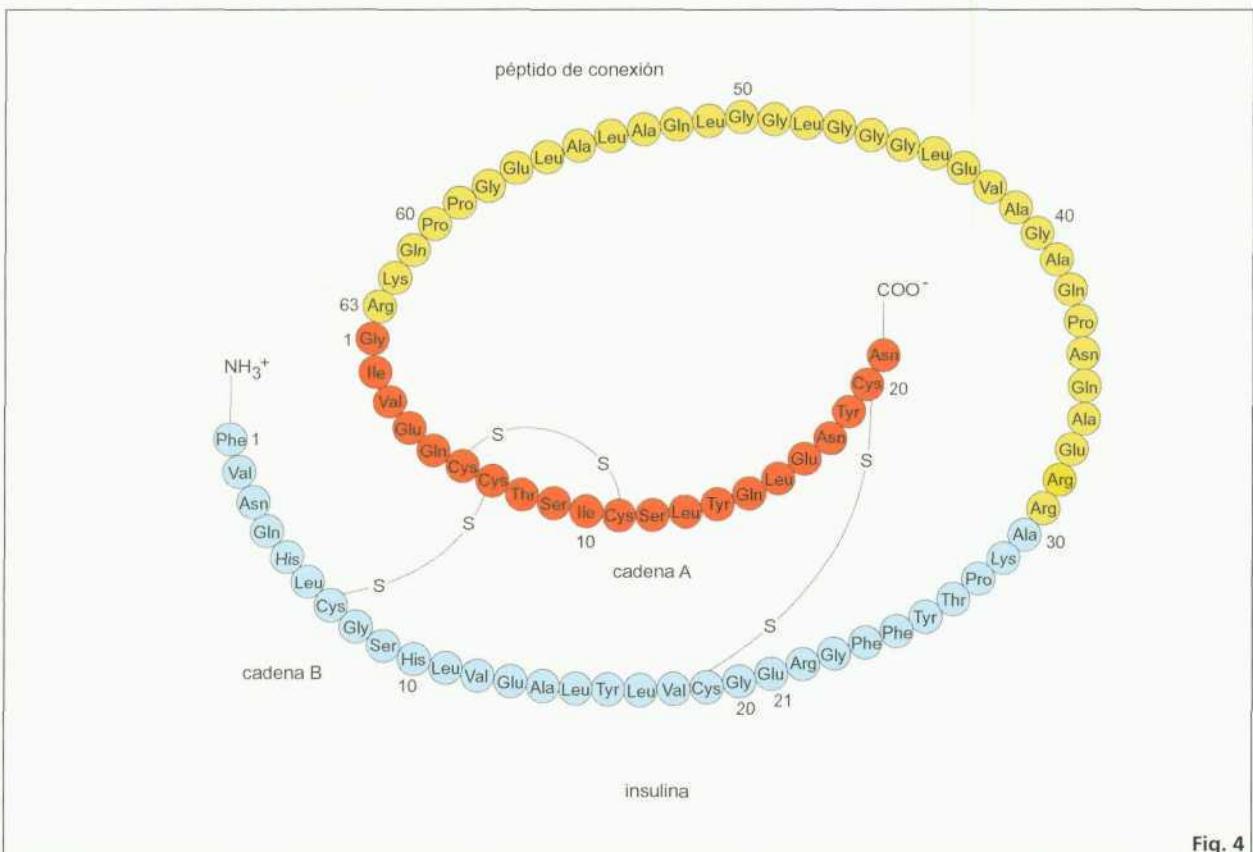
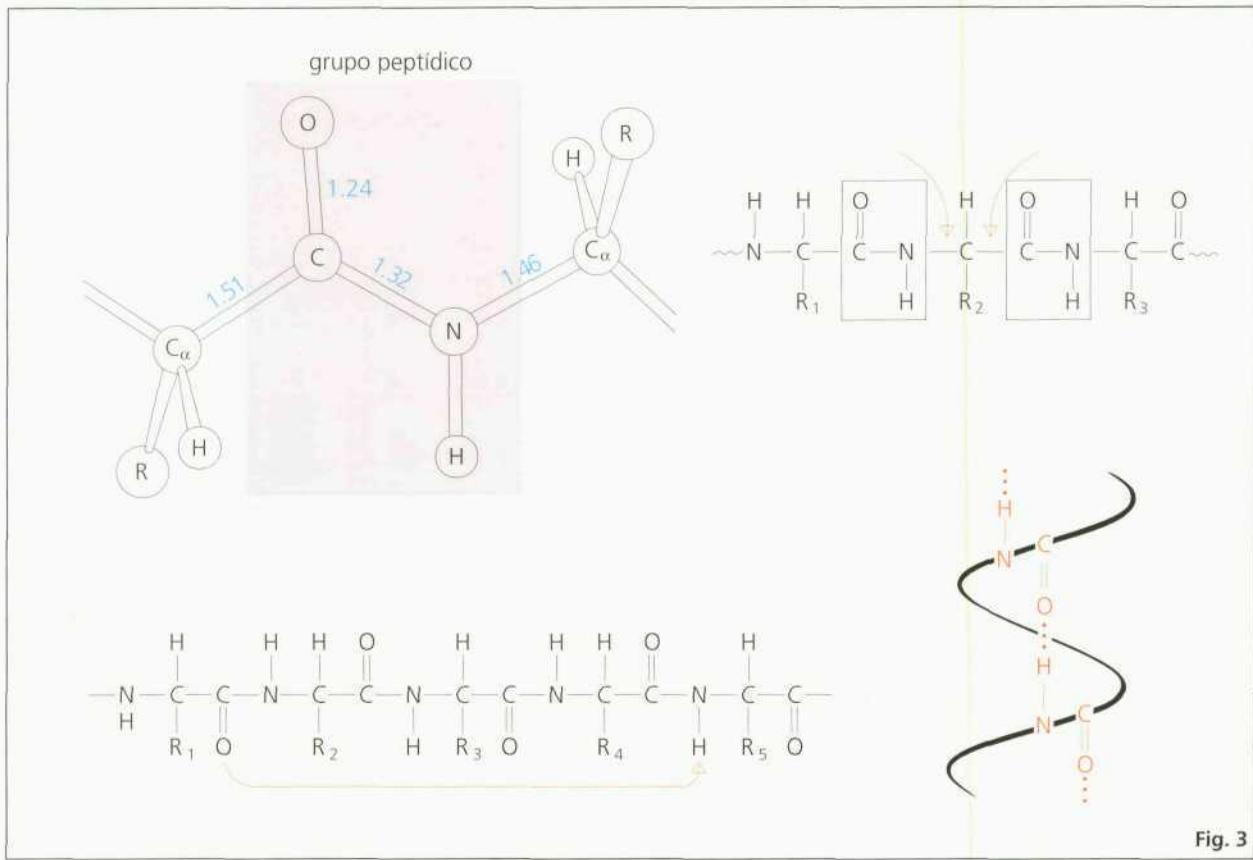


Fig. 2

pusieron a disposición de la bioquímica las herramientas más decisivas de su progreso en las últimas décadas. Entre otros, la *electroforesis en gel de agar* (1949), *en papel* (1950), *en almidón* (1952) y *en gel de poliacrilamida* (1959); el *electroenfoque* que aprovecha las diferencias en los puntos isoeléctricos de los componentes de una mezcla para lograr su separación en un gradiente estable de pH; y la *isotacoforesis* que aprovecha el gradiente de valores de pK, para, a pH constante, lograr un gradiente de movilidades electroforéticas. Efectivamente, las técnicas electroforéticas han constituido herramientas fundamentales en la investigación bioquímica para la separación de proteínas muy semejantes, como las isoenzimas de carga o las que resultan en los procesos de desamidación durante el envejecimiento.

miento. Los procedimientos cromatográficos tuvieron sus primeros fundamentos en la separación de moléculas, en particular de pigmentos vegetales, llevada a cabo hace una centuria por el ruso Tsvett utilizando lechos de óxido de aluminio o de carbonato cálcico. Estos trabajos permanecieron olvidados hasta que, tras un cuarto de siglo, grandes químicos como Willstätter, Kuhn y Lederer reutilizaron la misma técnica. En la década de los cuarenta, Martin y Synge desarrollaron la *cromatografía sobre papel*; y Martin y James, en 1952, la *cromatografía de gases*. Por los mismos años, Moore y Stein desarrollaron la *cromatografía de cambio de ión*, procedimiento extraordinario para la separación y cuantificación de aminoácidos en sus mezclas, principalmente en hidrolizados de proteínas.



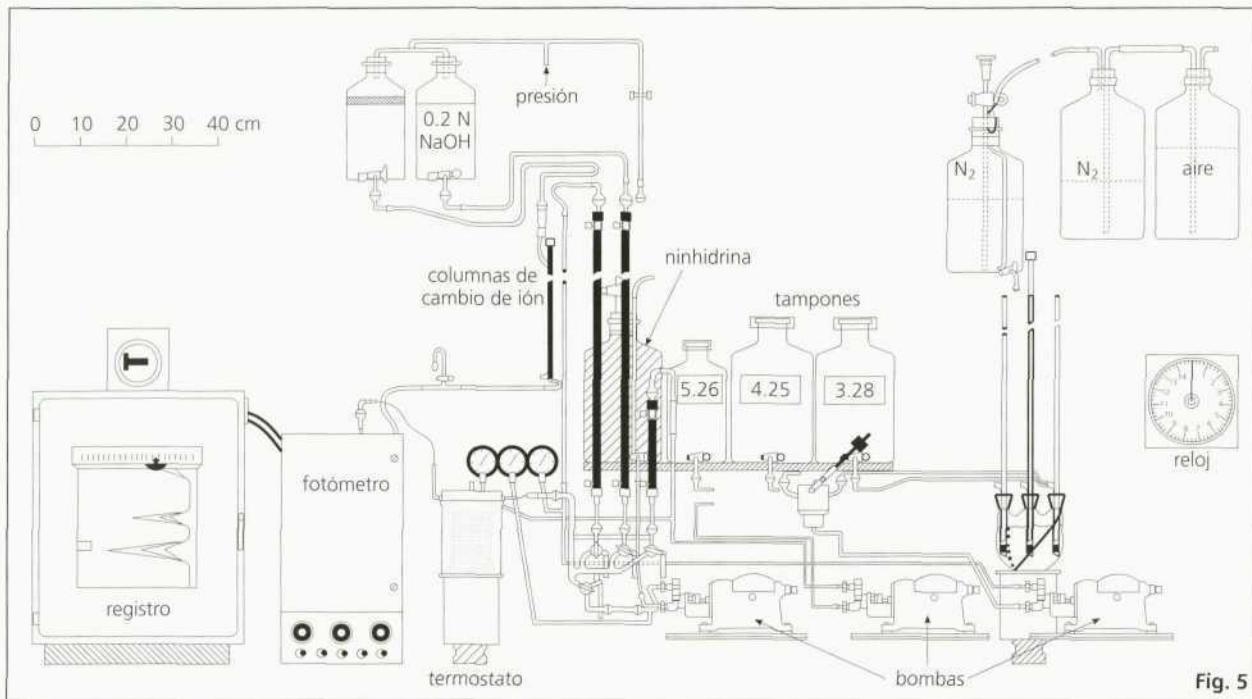


Fig. 5

Este procedimiento de separación, y su ulterior cuantificación por la reacción coloreada con ninhidrina, ha constituido la base de los *analizadores automáticos de aminoácidos*, cuyo esquema aparece en la figura 5. El uso de los analizadores de aminoácidos y el cálculo de masas moleculares por ultracentrifugación permitieron conocer la fórmula empírica de las proteínas, lo que significó indudablemente un cierto avance; pero, a diferencia de lo que ocurre con la generalidad de las moléculas, este conocimiento es poco relevante para penetrar en la compleja estructura de las proteínas y la averiguación de sus distintos niveles estructurales.

La denominada *estructura primaria* se refiere a la ordenación secuencial que los aminoácidos presentan a lo largo de la cadena polipeptídica; lo que constituye, como ya ha quedado representado en la figura 1, la *traducción* de la información contenida en los mRNA, soportada por una secuencia de bases, en la *secuencia de los aminoácidos*, específica para cada molécula de proteínas. Esta especificidad es doble; de un lado, la debida a la actividad biológica particular de la molécula –en orden a la *catálisis*, la *defensa*, el *transporte* de pequeñas moléculas, la variada acción de los distintos tipos de *receptores*, los mecanismos de la *transducción de señales*, incluido el fenómeno de *apoptosis*, etc.–; y, de otro, la propia especificidad de una especie biológica determinada, de forma que la insulina del hombre, por ejemplo, difiere en su estructura primaria de la de otras especies, como el caballo, el cerdo, etc. El conocimiento de esta diferencia entre las especies es esencial; en ella radica la mayor o menor facilidad con que una determinada proteína no humana –podemos seguir con el mismo ejemplo de la insulina– puede utilizarse, en función de su respuesta inmunitaria, en la terapéutica hu-

mana y, por ende, la importancia de la fabricación biotecnológica de las proteínas humanas por bacterias transgénicas.

Efectivamente, en 1945, Frederick Sanger inició un largo estudio que, en ocho años, le habría de conducir al conocimiento de la secuencia de aminoácidos –la *estructura primaria*– de la insulina. A dos circunstancias principales se debió el éxito de este trabajo. En primer lugar, a que la insulina carece de dos aminoácidos lábiles, la metionina y el triptófano, cuya presencia dificulta –lo hacía, sobre todo, en aquella época–, el tratamiento del conjunto; en segundo, a la puesta a punto de reactivos químicos –sobre todo, el fluordinitrobenceno, el cloruro de dansilo y el cloruro de dabsilo– (figura 6) que, aunque ya en desuso, son capaces de bloquear de manera bastante selectiva la posición  $-NH_2$  terminal de la molécula. A la determinación de la estructura primaria de la insulina siguieron las determinaciones de la *ribonucleasa* con 124 aminoácidos en una sola cadena, del *virus del mosaico del tabaco*, de la *tripsina*, y un largo etcétera que hoy abarca a muchísimos cientos de cadenas polipeptídicas.

Sin embargo, las condiciones químicas se vieron mejoradas enseguida por un nuevo reactivo que significó un extraordinario avance, el isotiocianato de fenilo,  $S=C=N-C_6H_5$ , descrito por Edman en 1950 (figura 7). El mismo autor describió, diecisiete años después, la automatización de la determinación de la secuencia de aminoácidos en fase líquida (figura 8), a cuya técnica han seguido nuevos sistemas que operan en fase sólida o, más recientemente, en fase gaseosa. En el *secuenciador* que aparece en la figura 8, la proteína objeto de estudio se coloca sobre las paredes de una cámara termostatizada de reacción por efecto de un campo centrífugo, en la que tiene

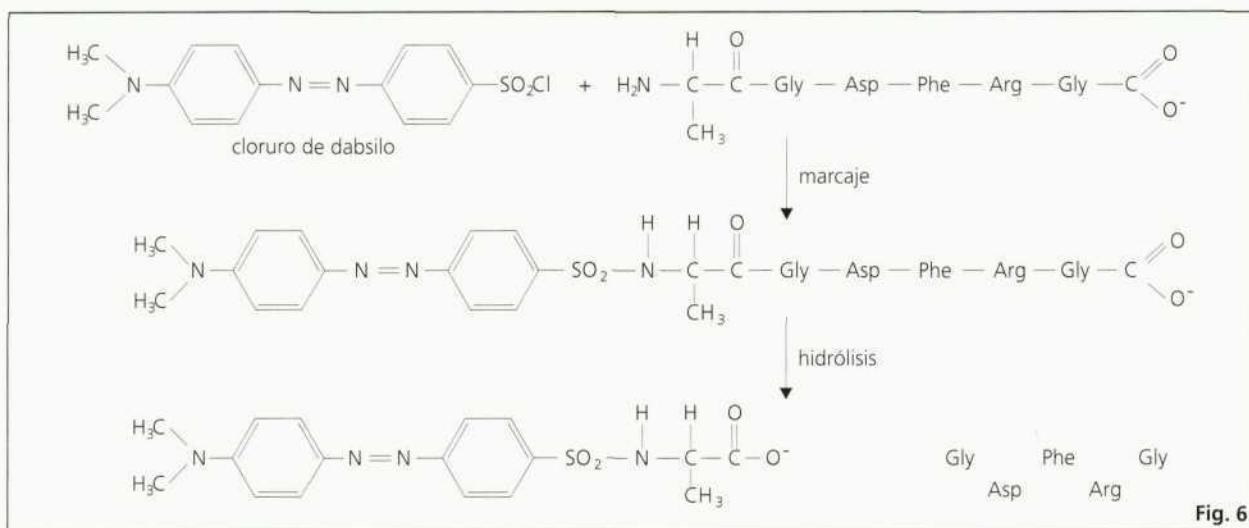


Fig. 6

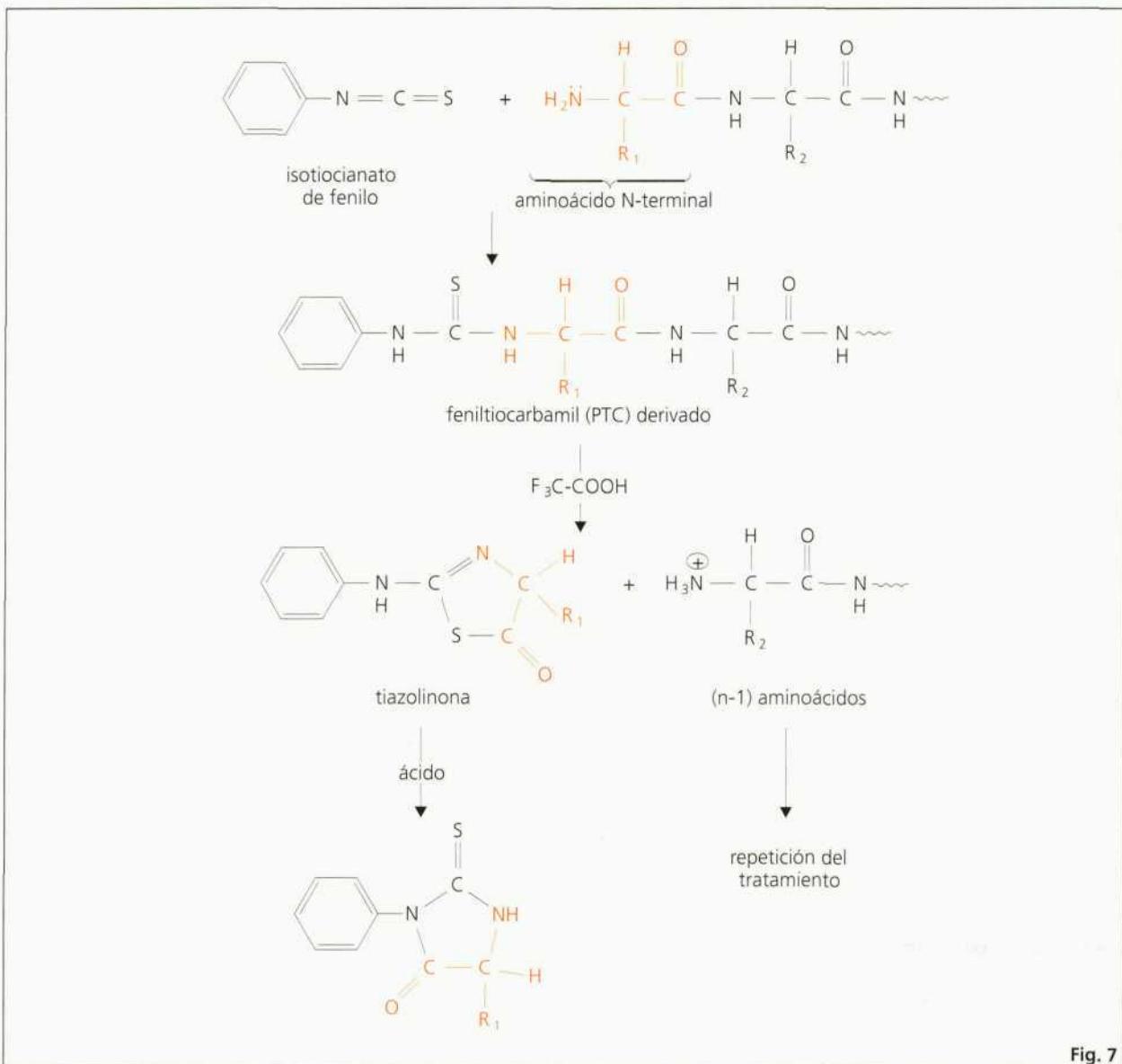


Fig. 7

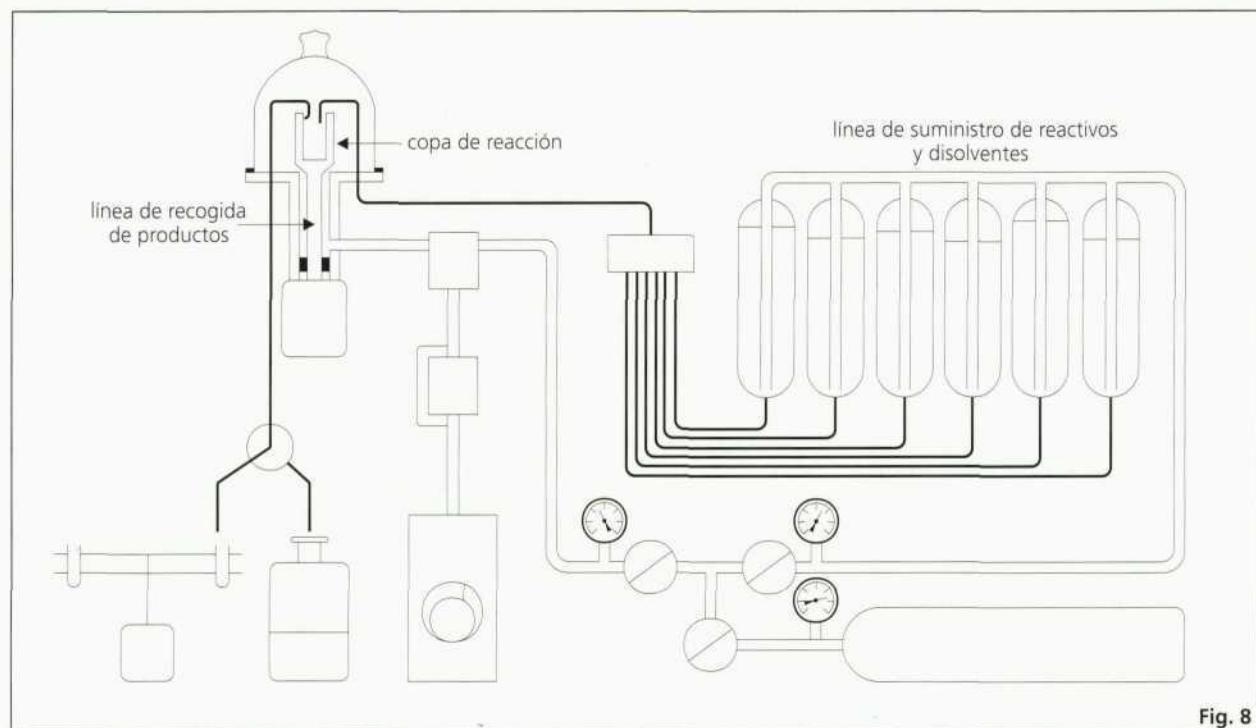


Fig. 8

lugar la adición y liberación programada de reactivos, productos y disolventes en una combinación de períodos de vacío y sobrepresión de nitrógeno. Con este tipo de aparatos, y debido al alto rendimiento de la reacción de Edman, se ha llegado a realizar más de un centenar de ciclos de degradación secuencial de una cadena polipeptídica. Y, cuando las proteínas son de gran tamaño, la determinación de su estructura primaria requiere varias etapas complementarias, como pueden ser la *rotura de los puentes disulfuro*, la rotura de cadenas por la *acción de reactivos específicos* como el bromuro de cianógeno sobre los residuos de metionina y la *acción de enzimas proteolíticas* con actividades específicas de rotura –pepsina, tripsina, quimotripsina, termolisinasa, principalmente–, para lograr fragmentos capaces de ser sometidos más eficazmente a la acción del secuenciador automático. Su mayor limitación estriba en la solubilidad de algunos péptidos, sobre todo de los de bajo peso molecular, que se arrastran excesivamente por los disolventes, con la consiguiente reducción del rendimiento. Este inconveniente se ha salvado con el desarrollo de la *degradación de Edman en fase sólida*, en la que los péptidos, cualquiera que sea su tamaño o solubilidad, se fijan de forma covalente a soportes insolubles, y la vasija de reacción se sustituye por una columna que contiene el problema unido al soporte. Más recientemente se ha descrito un procedimiento que combina la degradación automática en fase sólida con el empleo de reactivos en fase gaseosa para el acoplamiento y posterior ruptura; en él, la proteína o péptido se insolubiliza sobre la matriz de un polímero que se sitúa sobre una placa porosa de teflón, a través de la que circulan reactivos y disolventes. Frente a los dos sistemas iniciales presenta ciertas

ventajas: frente al sistema en fase líquida, la de utilizar cantidades mucho más pequeñas –del orden de 5 pmoles en lugar de los 5 nmoles del anterior–, y frente al sistema en fase sólida, la de no necesitar la unión covalente de la cadena peptídica, con la consiguiente mejora del rendimiento. Mediante el empleo de estas técnicas, los ocho años que duró el trabajo de Sanger para la determinación de la estructura primaria de la insulina hubieran quedado reducidos a unos cuantos días.

El conocimiento de la *estructura primaria* de las proteínas es fundamental para penetrar en los mecanismos de su *función* y, en consecuencia, para el establecimiento de las relaciones *estructura-función*. Porque en la estructura primaria descansa no sólo la *especificidad funcional* de las proteínas, como ha quedado subrayado, sino las posibilidades de sus *alteraciones patológicas* de índole genética e, incluso, el origen de sus *niveles superiores de estructura*. Y, aunque sea adelantarse un tanto en este estudio, habrá que señalar ahora que esta secuencia impone las condiciones de plegamiento de las cadenas para el logro de las conformaciones nativas propias de la estructura terciaria.

### SÍNTESIS DE PÉPTIDOS

La exactitud de los análisis estructurales tendría que verse confirmada por la correspondiente síntesis peptídica y la identidad del producto obtenido con el analizado. Ya en 1901, Fischer había iniciado los métodos químicos de síntesis de péptidos por medio de la activación química de los carboxilos de los aminoácidos, utilizando los esteres, los cloruros o las azidas de ácido. Con estos procedimientos

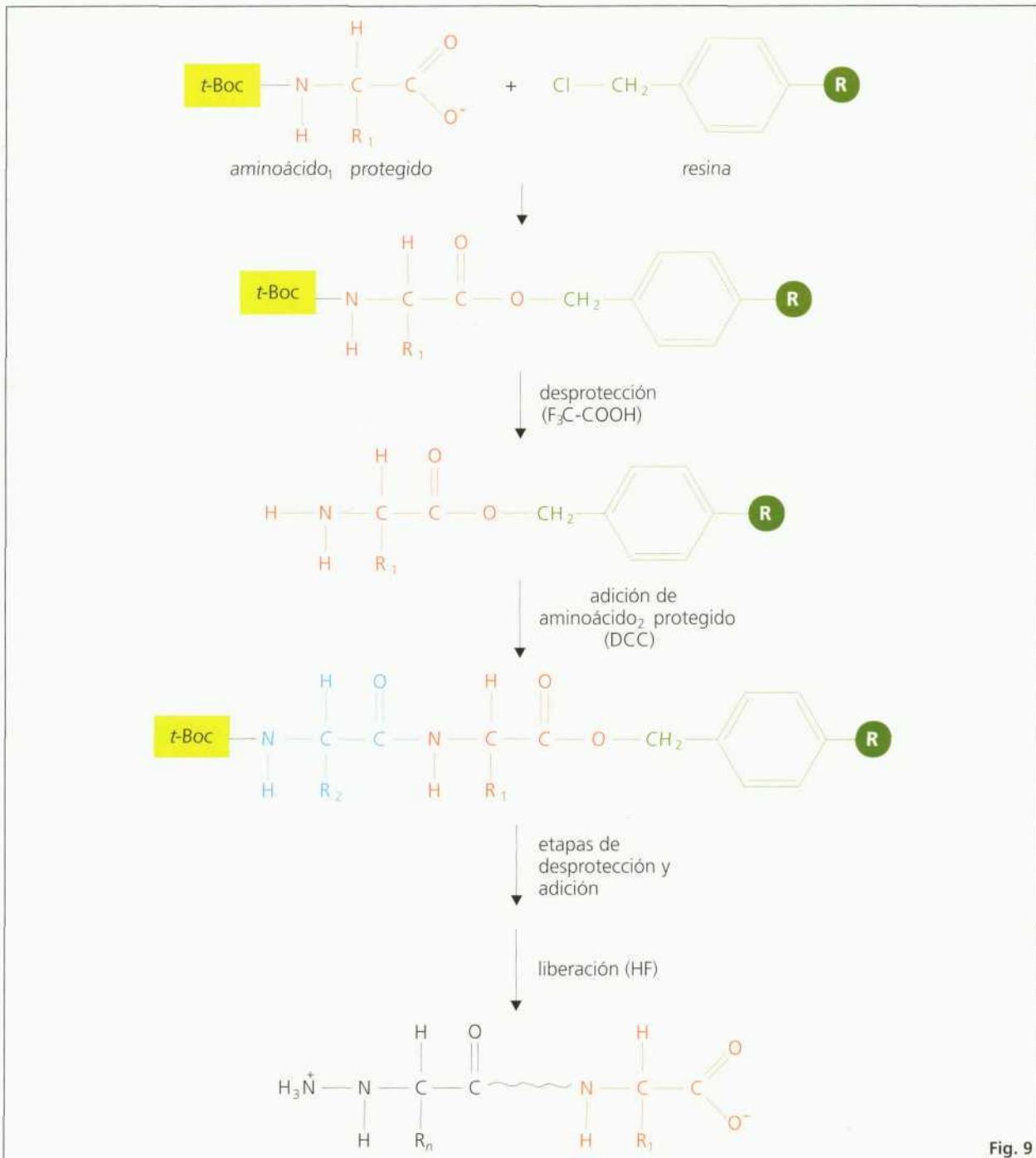


Fig. 9

clásicos de la química, logró Abderhalden, en 1916, sintetizar un nonadecapéptido. Sin embargo, la era moderna de la síntesis de péptidos no comenzó hasta 1932, cuando Bergmann y Zervas introdujeron un método singular de protección del grupo  $-\text{NH}_2$  de los aminoácidos por medio del cloruro de benciloxicarbonilo,  $\text{C}_6\text{H}_5-\text{CH}_2-\text{O}-\text{COCl}$ , que más tarde se amplió a la utilización del bloqueo por el *terbutil-oxicarbonilo* (*t*-BOC). La posterior activación del grupo carboxilo mediante la formación de anhídridos mixtos o el empleo de la diciclohexil-carbo-

diimida han constituido variantes con las que se llevó a cabo la síntesis de péptidos hasta que, en la década de los sesenta, Merrifield desarrolló la *síntesis de péptidos en fase sólida*. En este método (figura 9), el primer aminoácido, adecuadamente protegido en su grupo  $-\text{NH}_2$ , se une a una resina —constituida por un copolímero de 98 % de estireno y 2 % de polivinilbenceno, con los anillos aromáticos parcialmente clorometilados para originar el sitio de anclaje—, de forma que el posterior desbloqueo del  $-\text{NH}_2$  del aminoácido inicial permite la entrada del segundo aminoáci-

do igualmente protegido y en presencia de un agente de condensación como la diciclohexil-carbodiimida. Este conocimiento básico dio lugar al desarrollo del *sintetizador automático de péptidos*, con el que, en los años siguientes, se describió la síntesis de la ribonucleasa A (Merrifield), el citocromo c (Sano y Kurihara), la ferredoxina (Bayer, Jung y Hagen) y la hormona de crecimiento (Li y Yamashiro). Ha sido en la síntesis de péptidos de bajo peso molecular de aplicación farmacológica, o en la producción de regiones específicas con actividad antigenética para la producción de anticuerpos específicos y vacunas, donde el *sintetizador* ha alcanzado su máxima eficacia.

#### ESTRUCTURAS SECUNDARIA Y TERCIARIA

Hay que anticipar que los biopolímeros en general y las proteínas en particular deben su funcionalidad a un con-

junto de interacciones no covalentes condicionadas por la singular disposición espacial de su estructura química. Y la difracción de rayos X comenzó a resultar útil para el análisis estructural cuando, a comienzos del siglo XX, Von Laue descubrió que las distancias interatómicas de las redes cristalinas eran del orden de la longitud de onda de los rayos X, y podían servir para experimentar los fenómenos de difracción, de cuya interpretación habrían de surgir los datos acerca de las distribuciones atómicas de las redes.

En efecto, a finales de los años treinta, Linus Pauling y Robert Corey, en el Instituto Tecnológico de California, comenzaron un estudio sistemático de *difracción de rayos X* con aminoácidos, péptidos y proteínas, esencialmente de naturaleza fibrosa; de manera que en 1951 presentaron datos relativos a las *distancias y ángulos de enlace* de las cadenas polipeptídicas y elaboraron un modelo de *estructura helicoidal*, cuya característica principal era la de contener 3.6 residuos de aminoácido por cada giro (figura 10).

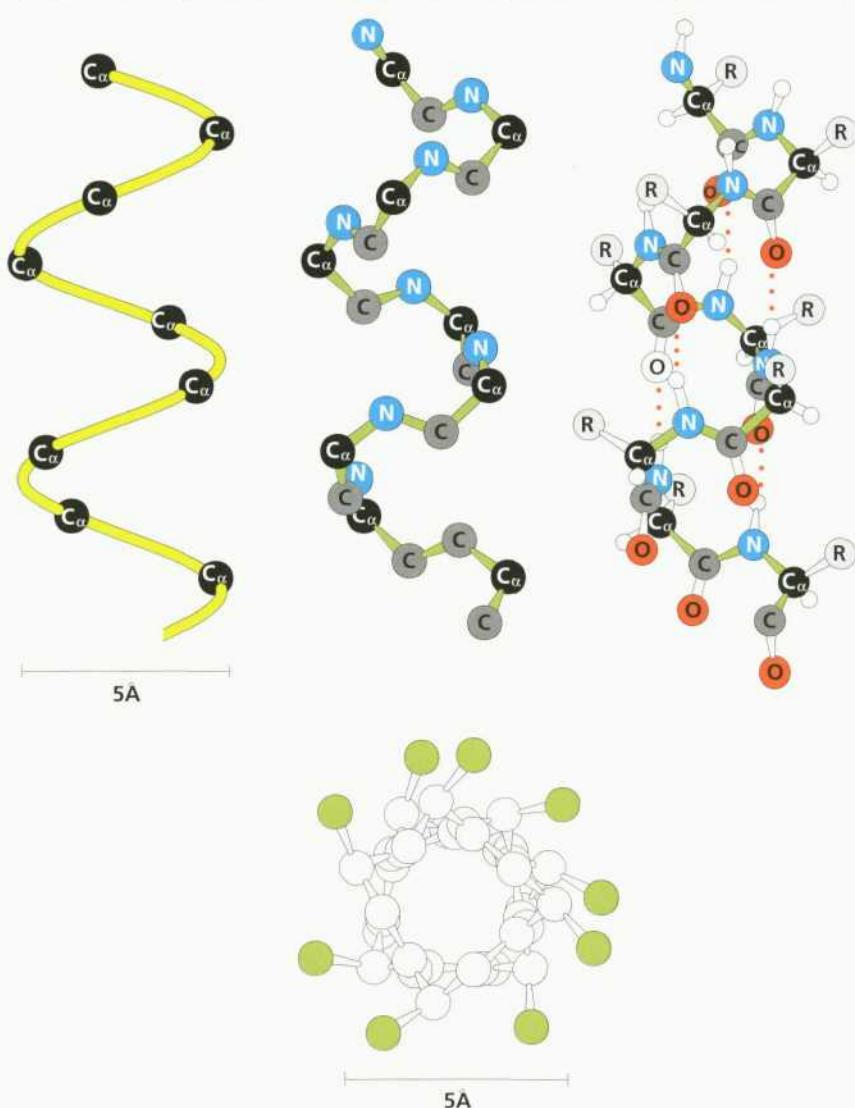
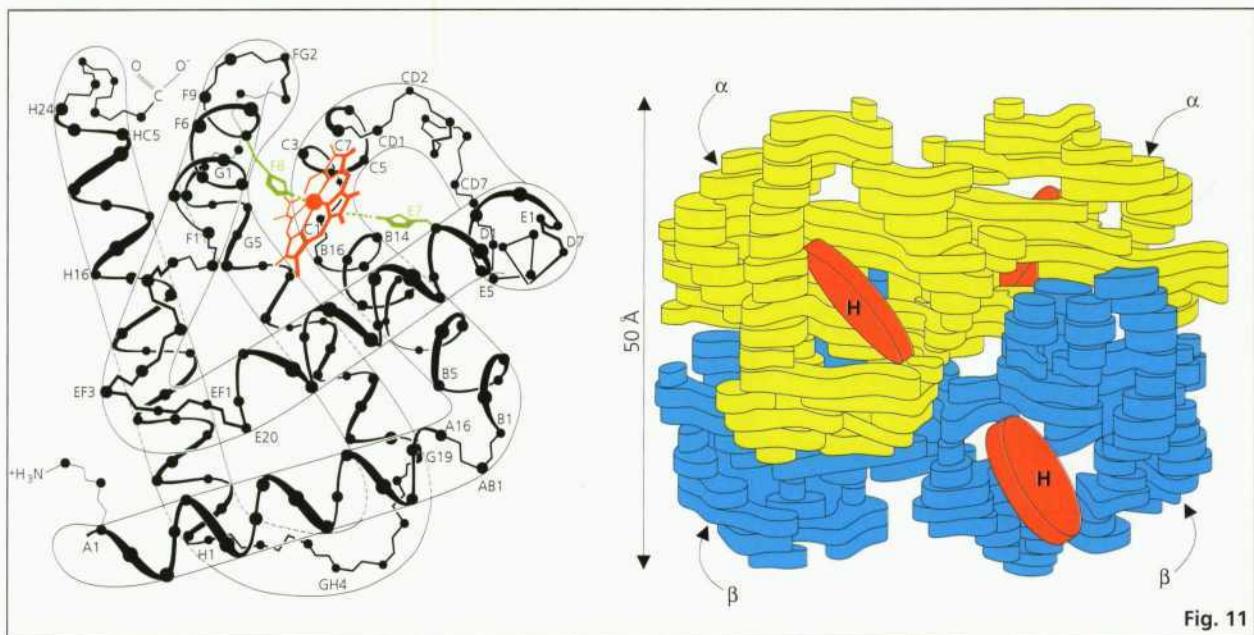


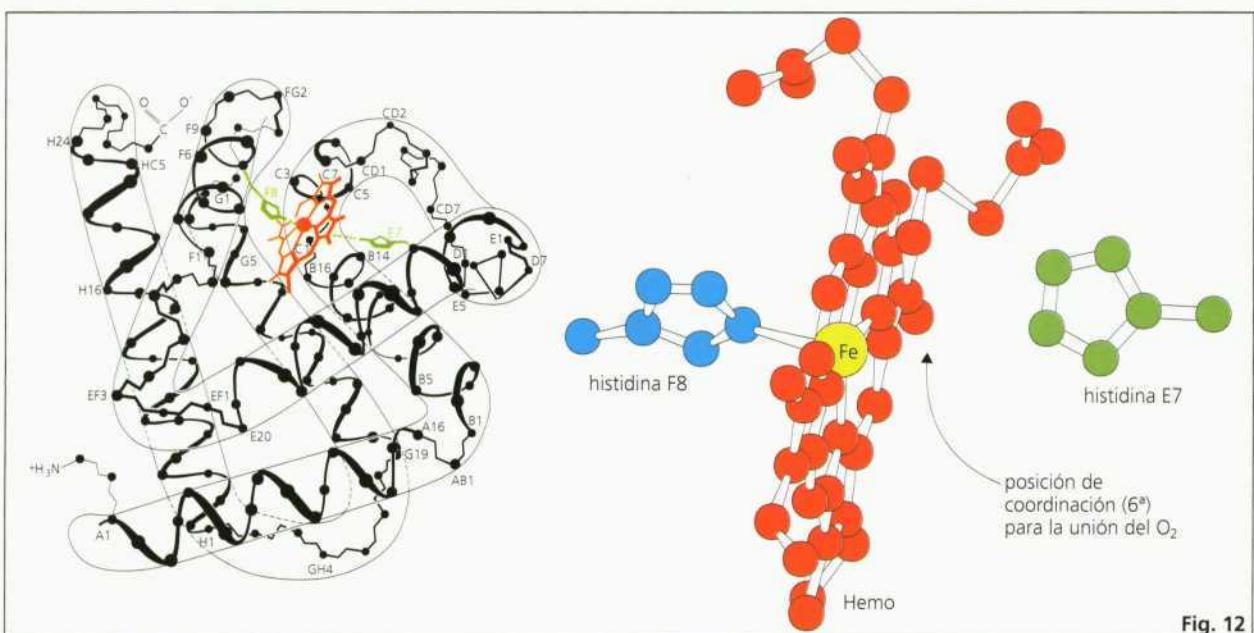
Fig. 10



Esta estructura de  $\alpha$ -hélice resultó ser uno de los plegamientos más habituales en la estructura de las proteínas. Esta disposición relativa en el espacio de dos aminoácidos consecutivos se conoce como *estructura secundaria* o, si se quiere, *repetición regular de ciertas conformaciones en el esqueleto peptídico*. Por la misma época, Max Perutz y John Kendrew dirigieron sus esfuerzos hacia la comprensión de los datos de difracción de proteínas globulares cristalizables, como fueron los casos de la *hemoglobina* y la *mio-globina*, las dos primeras proteínas globulares cuya estructura atómica detallada se pudo llegar a conocer (figura 11). Esta disposición molecular en el espacio en la que se establece la relación entre dos aminoácidos no con-

secutivos es la que se conoce como *estructura terciaria*.

A partir de entonces han sido numerosísimas las estructuras terciarias de proteínas identificadas a través de los esquemas de difracción de rayos X. Es el caso de diversas enzimas (lisozima, prostaglandina sintasa y otras sintetas, hidroxilasas, deshidrogenasas, fosfolipasas, adenilato ciclase, proteína quinasas, citocromo reductasas, caspasas, etc.), la albúmina sérica, los transportadores de oxígeno, las inmunoglobulinas y sus fragmentos, los factores de crecimiento, los receptores de los más variados ligandos (receptores de células T-MHC, de hormonas, etc.), los canales de iones, la bacteriorrodopsina, las moléculas implicadas en la transducción de señales, las proteínas y complejos fun-



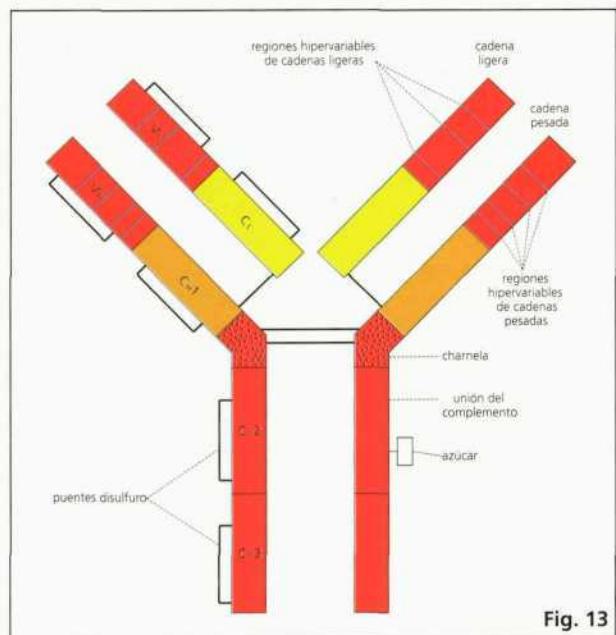


Fig. 13

cionales de los ribosomas, las citoquinas, las tubulinas, las integrinas, las dinaminas, las cadherinas y la maquinaria mitótica.

Más recientemente, la espectroscopia de RMN está contribuyendo eficazmente a la determinación estructural de las proteínas.

La asociación, por lo general no covalente, de entidades de proteínas con sus niveles estructurales respectivos, *primario, secundario y terciario*, origina las llamadas *estructuras cuaternarias*, de las que existen numerosos ejemplos de diferente complejidad.

A modo de ejemplo de estas *estructuras cuaternarias* podemos citar: la *hemoglobina* ( $\alpha_2\beta_2$ ), a base de cuatro sub-

unidades –dos subunidades  $\alpha$  y otras dos subunidades  $\beta$ – con sus correspondientes niveles estructurales, cada una de las cuales interacciona con un anillo porfirínico que contiene Fe para el transporte y cesión del O<sub>2</sub> con arreglo a un mecanismo singular (figura 12); la *inmunoglobulina G*, que cumple sus funciones con unas estructuras singulares, entre otras las de la *región charnela* y las zonas *hipervariables*, responsables de la variabilidad de la respuesta antigenica en el seno de una original estructura a base de dos cadenas pesadas y dos cadenas ligeras, con puentes disulfuro intra e inter-catenarios (figura 13); y las proteínas de las clases I y II del *complejo principal de histocompatibilidad*, cuyas múltiples zonas responden a la necesidad de las interacciones moleculares en la presentación de antígenos a las células T (figura 14).

En la actualidad, los bancos de datos sobre proteínas recogen la información sobre el tamaño y la forma de más de un millar de moléculas.

#### TAMAÑO, FORMA Y FUNCIÓN. NUEVAS ESTRUCTURAS Y NUEVAS FUNCIONES

Proteínas son los principales catalizadores, elementos estructurales, máquinas moleculares y transductores de señales biológicas (tabla IV). La gran singularidad estructural de las proteínas radica en la multiplicidad de sus niveles estructurales y en la extraordinaria variabilidad de cada uno de ellos; y como consecuencia de todo ello resulta un muestuario formidable de *tamaños y formas*, cuya relación con la función biológica resulta hoy difícil de evaluar.

Hasta época reciente existían dos métodos principales para profundizar en la función de las moléculas de las proteínas. De un lado, el conocimiento primario de la fun-

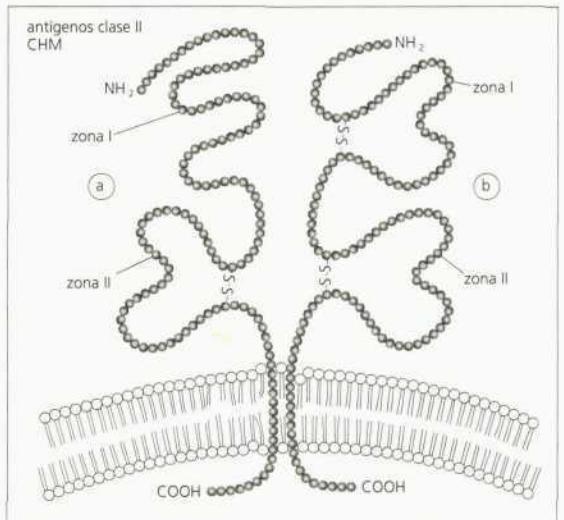
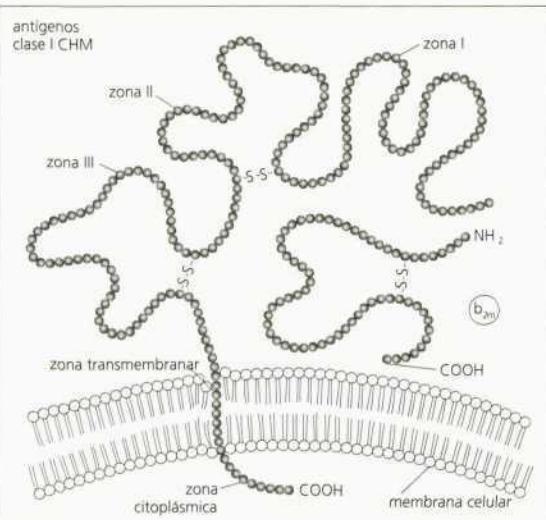


Fig. 14

**Tabla IV. Funciones de las proteínas**

<b>1. Catálisis enzimática:</b>
Las enzimas incrementan extraordinariamente las velocidades de reacción. varios millares han sido caracterizadas y muchas han sido cristalizadas.
<b>2. Transporte y almacenamiento:</b>
Muchas moléculas pequeñas –O <sub>2</sub> , Fe, por ejemplo– se transportan por proteínas específicas: hemoglobina, mioglobina, transferrina.
<b>3. Soporte mecánico:</b>
Las proteínas fibrosas como el colágeno, contribuyen al mantenimiento de las propiedades mecánicas de los huesos y la piel.
<b>4. Receptores de membrana:</b>
La respuesta a la acción de factores externos a la célula –luz, hormonas, neurotransmisores, opiáceos, toxinas, factores de crecimiento, etc.– se origina en su interacción con proteínas.
<b>5. Factores de crecimiento y diferenciación:</b>
Gran número de proteínas cumplen esta función en células específicas –plaquetas, fibroblastos, neuronas, células epiteliales–. El control de la división celular y de la respuesta inmunológica se realiza por medio de proteínas específicas.
<b>6. Anticuerpos:</b>
Estas proteínas –las inmunoglobulinas– forman parte de los sistemas de defensa.
<b>7. Coordinación de movimientos:</b>
Las proteínas participan en la contracción muscular, movimiento de cromosomas, microtúbulos, flagelos.

ción ha resultado de las observaciones bioquímicas, genéticas y estructurales de una proteína individual. De otro, y una vez ha sido asignada una función determinada a una proteína individual, se puede buscar la relación de esta función con la de otras proteínas mediante el *método de homología*, con el que se investiga la semejanza de secuencias de aminoácidos con la secuencia de la proteína original. A través de los métodos de homología, entre el 40 y el 70 % de las nuevas secuencias genómicas pueden asignarse a alguna función, incluso de las definidas en las células procarióticas. Y en la actualidad se han desarrollado nuevos métodos para detectar las conexiones funcionales entre las proteínas. Dentro de ellos, los programas BLAST se utilizan para extender el conocimiento experimental de la *función de las proteínas* a las nuevas secuencias genómicas.

Al tradicional método de la *homología de secuencias de aminoácidos* se han añadido recientemente nuevos *métodos computacionales*, nacidos del conocimiento completo de las secuencias genómicas, en los que las correlaciones se establecen entre pares de proteínas heredadas en varias especies –método de los *perfils filogenéticos*–, entre dominios de proteínas que existen como consecuencia de la fusión o entre polipéptidos libres –método de la *piedra Rosetta*–, o entre la posición de los genes sobre los cromosomas –método del *gen vecino*–. Este tipo de análisis da lugar a complejas redes de conexiones funcionales entre las proteínas de las células y, como ha quedado mencionado, altera fundamentalmente la idea de la llamada *función de una proteína*.

Un *perfil filogenético* describe un esquema de relaciones acerca de la presencia o ausencia de una proteína particular en una serie de organismos de genomas conocidos. En concreto, este *método de los perfils filogenéticos* (figura 15) considera, de un lado, los genomas de una serie de espe-

cies (*A*, *B*, *C* y *D*) y, de otro, una colección de proteínas (*P*<sub>1</sub> a *P*<sub>7</sub>) que, por completo presentes en *A*, están presentes (1) o ausentes (0) en cada una de las otras tres especies, *B*, *C* y *D*. Los perfiles de cada una de las proteínas *P*<sub>1</sub> a *P*<sub>7</sub> se reúnen en conjuntos de forma que aparezcan reunidas las proteínas con los mismos perfiles –en este caso, el par *P*<sub>2</sub>-*P*<sub>7</sub> y el par *P*<sub>3</sub>-*P*<sub>6</sub>–, y que aparezcan conectados los perfiles con una sola distinción. La conclusión es que si dos proteínas poseen el mismo perfil filogenético –es decir, el mismo esquema de presencia o ausencia– en los genomas examinados, se infiere que ambas proteínas poseen una relación funcional; porque ¿cómo cabría pensar en una evolución a otras especies a no ser por esta relación funcional? No obstante, por lo general no se trata de proteínas homólogas, o, lo que es igual, esta relación funcional no requiere secuencias similares. La potencia de este método de los *perfils filogenéticos* crece con el número de posibles perfiles considerado; ya que cada proteína puede estar presente o ausente en cada genoma, y si el número de genomas secuenciados es *n*, el número de perfiles será  $2^n$ . Resulta, pues, que si pueden conocerse las secuencias de 30 genomas, el número de *perfils filogenéticos* posibles será  $2^{30} (\approx 10^9)$ . Número que excede el número de familias de proteínas, de forma que el *perfil filogenético* de una proteína podrá ser una forma de caracterización prácticamente única del esquema de distribución entre los genomas. Y de aquí que dos proteínas con idénticos o semejantes perfiles filogenéticos puedan considerarse implicadas en procesos o estructuras complejas comunes.

Así pues, el *perfil filogenético* de una proteína describe la presencia o ausencia de homólogos en los organismos. Las proteínas que forman parte de complejos multiméricos deberán poseer perfiles similares. Y, de igual manera, las proteínas que participen en determinados procesos metabólicos deberán ser vecinas en cuanto a sus perfiles filogenéticos. Estos hechos indican que la *comparación de perfils filogenéticos* no sólo es una herramienta de gran utilidad para la identificación de aquellas proteínas que tienen una presencia común en complejos estructurales o en la participación en procesos metabólicos, sino también para llevar a cabo asignaciones funcionales de proteínas no caracterizadas.

Con cierta frecuencia, dos proteínas independientes, *A* y *B*, de un organismo se expresan en otras especies bajo la forma de *proteínas fusionadas*. Y, en este caso, existen muchas posibilidades de que los dominios *A* y *B* posean relaciones funcionales. Dado, por otro lado, que *A* y *B* no poseen secuencias relacionadas, este tipo de conexión no puede detectarse a través de su homología y ha de hacerse a través de las secuencias genómicas. Este procedimiento se conoce como el *método de la piedra Rosetta*.

Un tercer método computacional para poner de manifiesto relaciones funcionales en las secuencias genómicas es el *método del gen vecino*. Si en diversos genomas los genes que codifican dos proteínas se encuentran vecinos en el cromosoma correspondiente, las proteínas tienden a encontrarse funcionalmente relacionadas. Dadas, sin em-

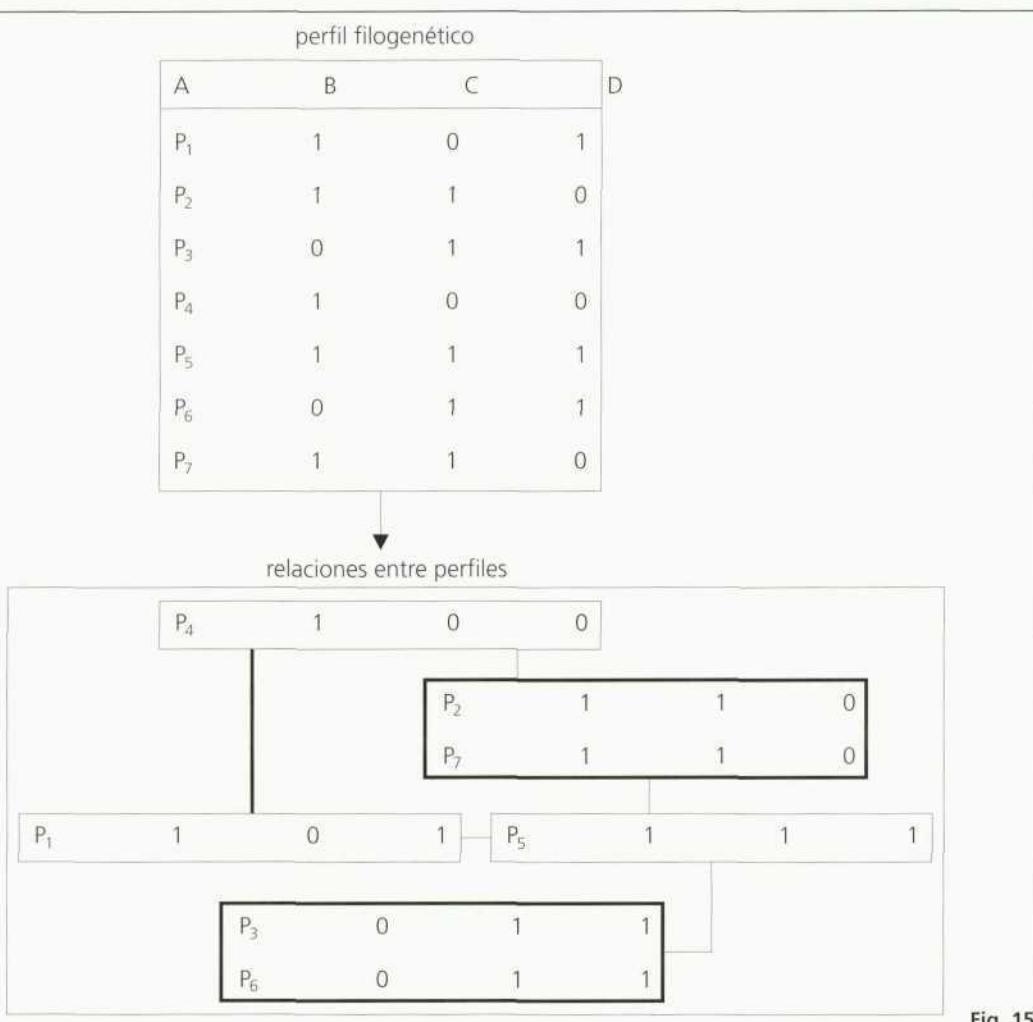


Fig. 15

bargo, las especiales características de la regulación genética procariótica –con la presencia de *operones* comunes–, este método es, de momento, más efectivo en este tipo de organismos.

En cuanto a la relación de la función con la estructura correspondiente, es obvio que existen estructuras peptídicas tan sencillas como las de los nonapéptidos hormonales de la glándula hipofisaria –oxitocina, vasopresina– y, frente a ellas, las de los grandes complejos de las deshidrogenasas, con más de 60 000 aminoácidos. De otro lado, y dentro de las enzimas mismas, existen moléculas formadas por una única cadena peptídica, como las hidrolasas, con alrededor de 125 aminoácidos, en tanto que la glutammina sintasa es un dodecámero con alrededor de 5 500 aminoácidos. En las figuras 16 y 17 se pueden ver los esquemas de una serie de proteínas con diferentes tamaños y formas, representados a escala, y responsables de una colección variable de actividades biológicas.

Indudablemente, el tamaño y la existencia de proteínas oligoméricas, así como los plegamientos de las cadenas individuales para dar lugar a diferentes dominios, contribuyen a la forma de las moléculas, vinculada íntimamente

a su función. Ya ha quedado mencionada la forma en Y de las inmunoglobulinas como respuesta a la necesaria flexibilidad, imprescindible a la capacidad de unión de ligandos como función primaria del sistema inmunitario. Las proteínas cuya actividad reside en su naturaleza reguladora del DNA adoptan formas con dominios capaces de adaptarse a los surcos del sustrato. Con frecuencia, los sitios activos de las proteínas oligoméricas se localizan en las interfaes entre las subunidades, permitiendo movimientos suaves entre ellas como ingredientes importantes de la catálisis o la regulación. En otras ocasiones, las proteínas oligoméricas posibilitan la canalización del sustrato hacia los sitios activos específicos, ganando en eficacia y especificidad, como sucede en el complejo de piruvato deshidrogenasa.

Dada esta frecuente asociación de unidades monoméricas, cabe preguntarse por los motivos de esta preferencia frente a las largas cadenas polipeptídicas con múltiples sitios activos. Y, entre las diversas interpretaciones de este hecho, cabe señalar, en primer lugar, la de que los errores de la biosíntesis se reducen en las cadenas de menor longitud. De otro lado, se ha señalado un aumento de

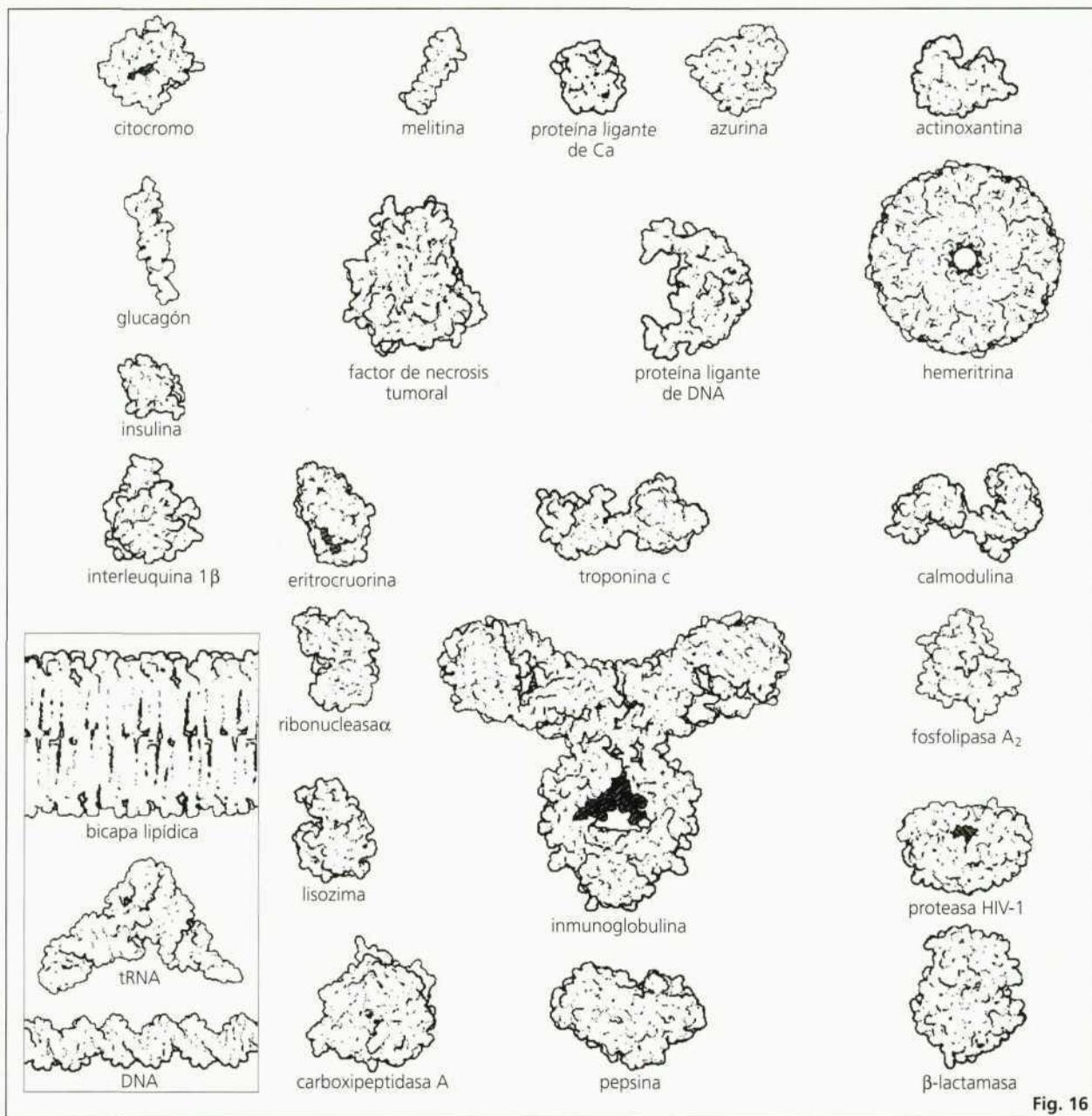


Fig. 16

la función a través de la asociación y disociación de las subunidades, con la posibilidad de eliminación de las subunidades erróneas y el aumento consiguiente de los oligómeros activos. De esta forma, cabe también la posibilidad de síntesis de los monómeros en un lugar, y, tras su difusión, su asociación como oligómeros funcionales en un sitio distante del anterior. O, como en el caso de la hemoglobina de la lamprea, la asociación y disociación completas de las subunidades median una transición cooperativa. Finalmente, se ha señalado que las proteínas oligoméricas son capaces de experimentar una selección más estricta en el curso de la evolución. En los multímeros homólogos se amplifica el efecto de las mutaciones; por ejemplo, la mutación de un residuo en la interfase entre las

subunidades de una proteína dimérica cambiará el carácter de los dos contactos. En tanto que, en los oligómeros superiores, el efecto se incrementa proporcionalmente. Y, a este propósito, se ha señalado que este hecho constituye una ventaja selectiva a lo largo de la evolución.

Además de estas características generales de tamaño y forma, existen algunas estructuras muy singulares que, lógicamente, influyen sobre su función. Entre otras, las *superhélices* o tirabuzones de  $\alpha$ -hélices, inicialmente descritas por Pauling y Corey en 1953, pero cuya resolución no ha sido lograda hasta los últimos años. Estas superhélices (*coiled coils*) constituyen importantes elementos estructurales (figura 18) de una abundante clase de proteínas fibrosas responsables de muy variadas funciones.

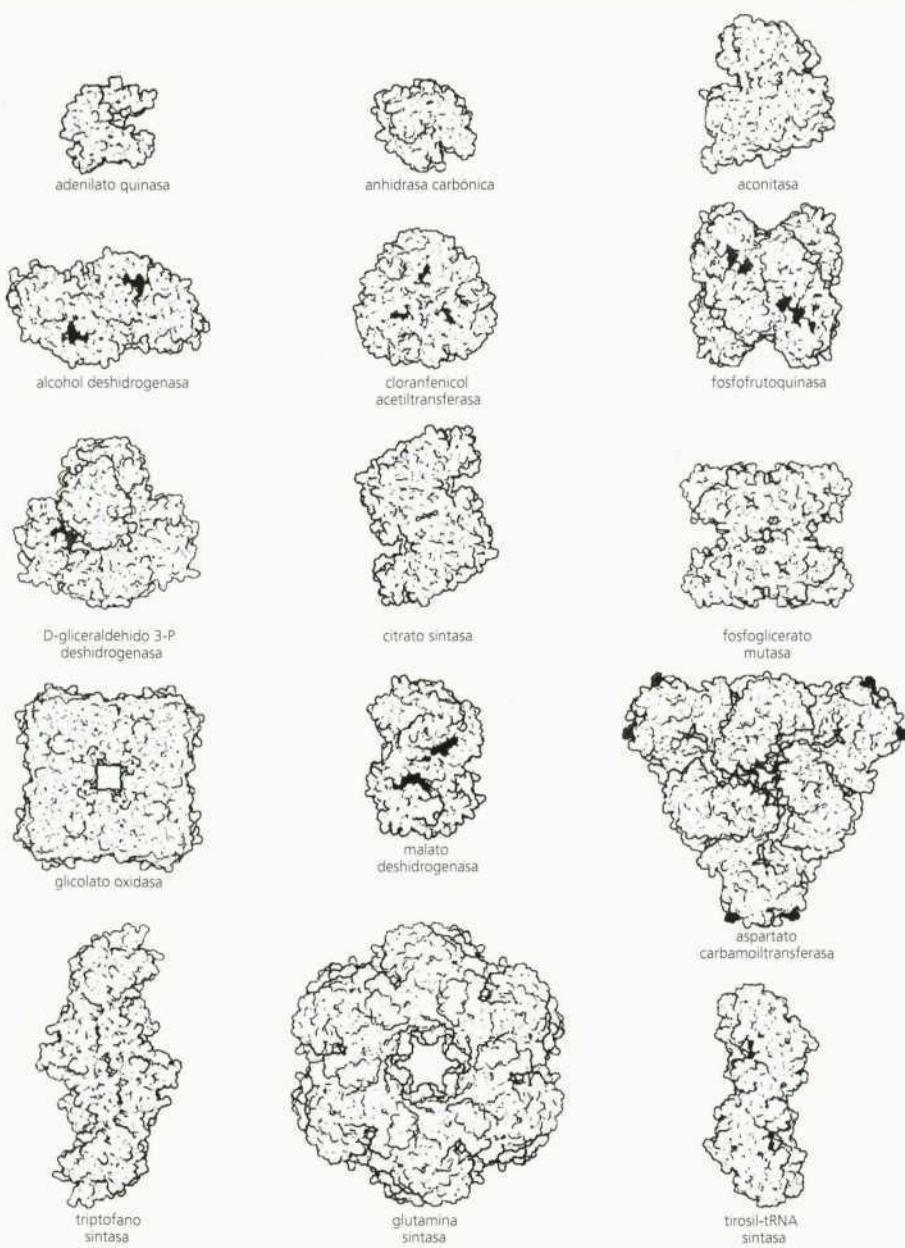
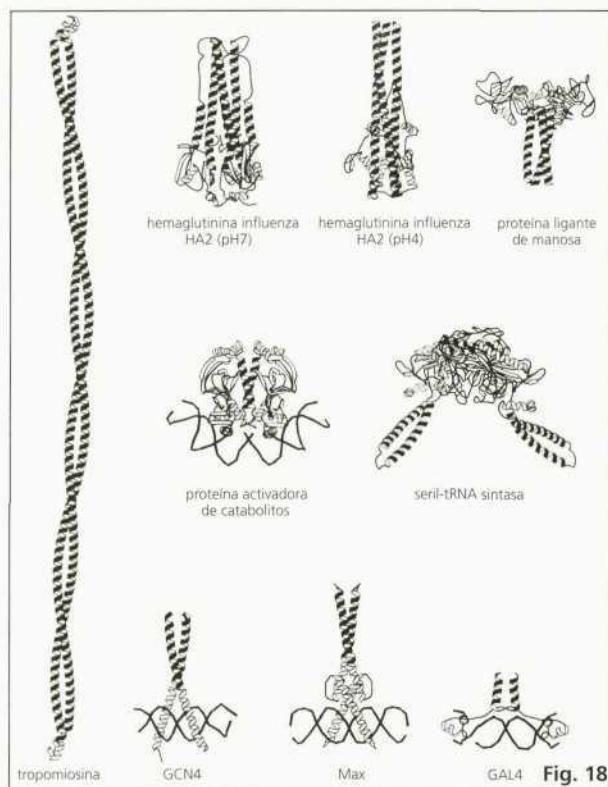


Fig. 17

Seguramente, su ejemplo más característico, por el carácter superhelicoidal casi exclusivo de la molécula, es el de la tropomiosina. Sin embargo, el interés por estas proteínas surgió al comienzo de la década de los ochenta al determinarse por alta resolución la estructura de los cristales de la *hemaglutinina* del virus de la gripe y observar que se trata de una superhélice de tres hebras cuya longitud varía con el pH. Una estructura semejante posee la *proteína ligante de manosa*. A continuación se fueron investigando otras proteínas en las que la superhélice ocupa posiciones muy variadas en el seno de la molécula. Así, por ejemplo, la superhélice se encuentra enterrada en la interfase del dímero que constituye la *proteína receptora de c-AMP o proteína activadora del gen*

*de los catabolitos*; en tanto que se encuentra libre prácticamente de interacciones con el resto de la molécula en la *seril-tRNA sintasa*. En 1988 se identificó su presencia como el elemento de dimerización en una clase de factores de transcripción, las llamadas *cremalleras de leucina (leucine zipper)*.

Dentro de estas *superhélices* existen numerosas posibilidades de variación de acuerdo con ciertas medidas físicas, tales como la longitud del paso de hélice y el ángulo de cruzamiento de las hélices (figura 19a). Otras posibilidades de variación resultan de la naturaleza «paralela» o «antiparalela» de las hélices y del distinto tipo de empaquetamiento de las cadenas laterales de los aminoácidos en el corazón del bucle de hélices. De esta



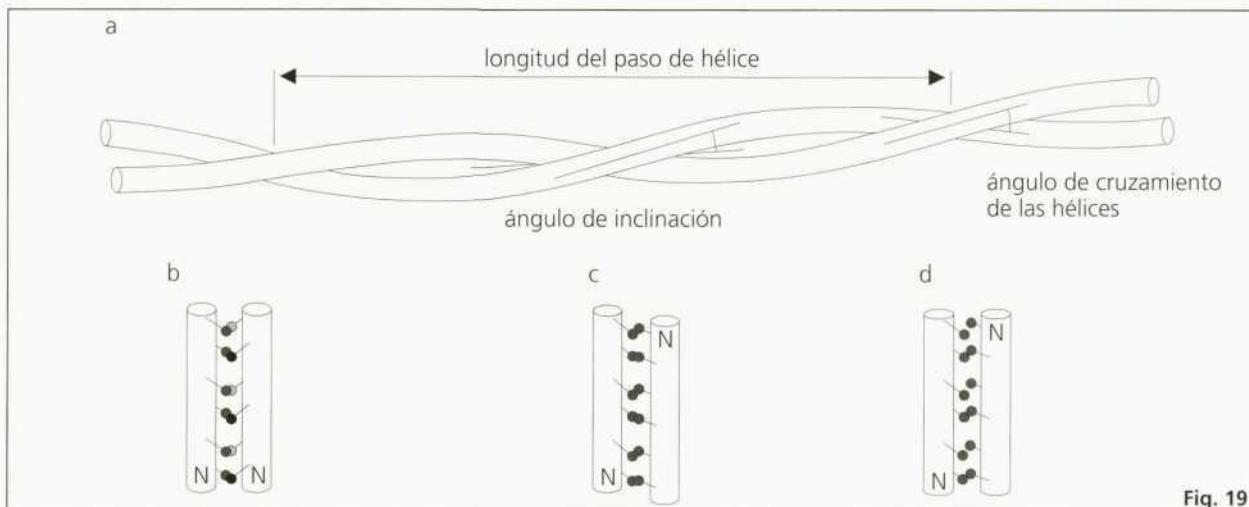
manera se producen dos tipos principales de estructuras: aquellas de los *nudos intracavitarios* (figuras 19b y 19c), en que un residuo de una hélice –el botón o nudo– se encuentra rodeado por cuatro cadenas laterales de la hélice con la que se enfrenta, y contactando con el residuo equivalente de la otra hélice (*knobs-into-holes*); y aquellas de *lomos acanalados* (figura 19d), más irregulares y características del empaquetamiento de las proteínas globulares helicoidales (*ridges-into-grooves*). Ambas estructuras pueden originar formas mixtas y, asimismo, presentar discontinuidades que producen distorsiones de las hélices. Las funciones de estas

superhélices han resultado ser extraordinariamente variadas: largas estructuras de gran resistencia mecánica como la *queratina* de pelos y plumas o la *fibrina* de los coágulos sanguíneos; las superficies protectoras de los patógenos, como las *proteínas M* de los estafilococos; las lipoproteínas que separan la membrana externa de la pared celular en las bacterias; suministran un esqueleto para los complejos reguladores, tipo *tropomiosina*; los soportes y levas moleculares, como las estructuras de *quinesina*, *hemagglutininas* y *miosina*; las estructuras dinámicas de varias familias de *activadores transcripcionales*; los «dedos» y «brazos» antiparalelos de *enzimas* que utilizan *DNA* y *RNA* como sustratos; en la *hemagglutinina* del virus de la gripe, la superhélice central suministra una especie de charnela que se pone en marcha tras una disminución del pH para impulsar una fusión peptídica hidrofóbica, que toma parte del mecanismo de liberación del virus en la célula, a través de una fusión de membrana.

#### UNIDADES MODULARES DE LAS PROTEÍNAS

Mediante los métodos de cristalografía de rayos X y de espectroscopía de resonancia magnética, se ha demostrado que muchas proteínas de los organismos multicelulares están construidas a partir de varios módulos o dominios autónomos, fácilmente identificables, entre las que se encuentran las proteínas responsables de la *coagulación*, la *fibrinólisis*, el *complemento*, la *matriz extracelular*, la *adhesión de la superficie celular* y los *receptores de citoquinas*. La figura 20 muestra representaciones de la estructura tridimensional de algunos de los módulos descritos. Las *proteínas modulares* participan especialmente en los mecanismos de *homeostasis sangre-coágulo*, la *adhesión celular* y la *transducción de señales*.

Hasta el momento no existen reglas definidas sobre el ensamblamiento de estos módulos en la construcción de las proteínas multimodulares. Por cristalografía se ha de-



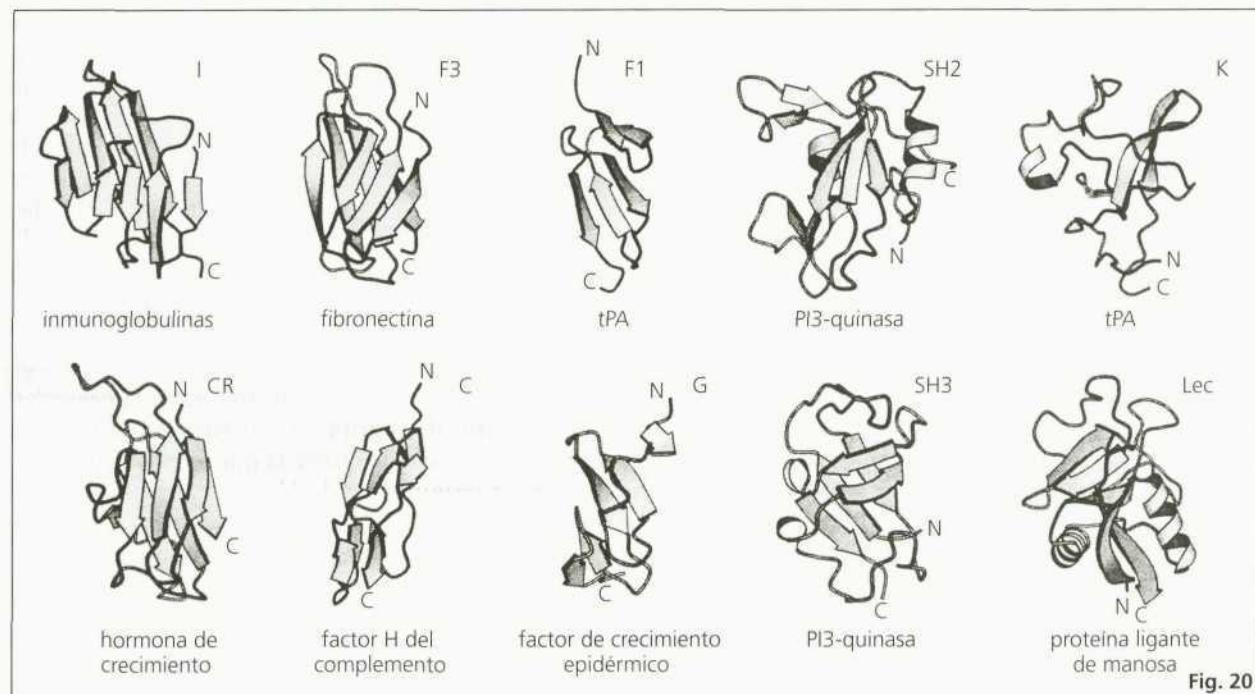


Fig. 20

mostrado la existencia de los pares I-I en numerosos casos, G-Lec en la *selectina* y F3-F3 en la molécula de adhesión *neuroglian*. Por resonancia magnética se ha probado la presencia del par F1-F1 en la *fibronectina*, C-C en el factor H del *complemento*, y el par F1-G en el *activador de plasminógeno tisular*. Un importante factor en la construcción modular de las proteínas es la relativa disposi-

ción de las posiciones N- y C-terminales y la longitud de las zonas intercalantes.

La figura 21 muestra un buen ejemplo de la utilización económica de estos módulos en la construcción de proteínas de superficie que median importantes interacciones proteína-proteína, como las participantes en los mecanismos biológicos propios de la presentación de antígenos a las células T.

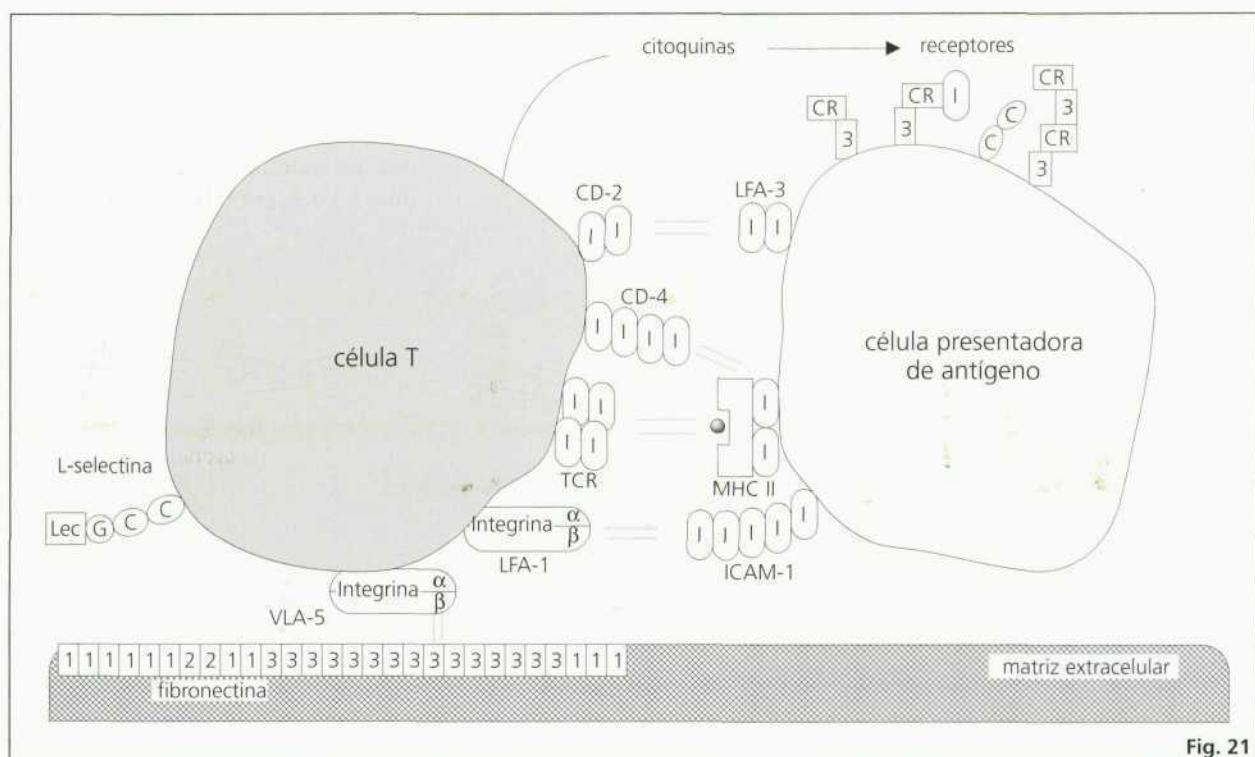
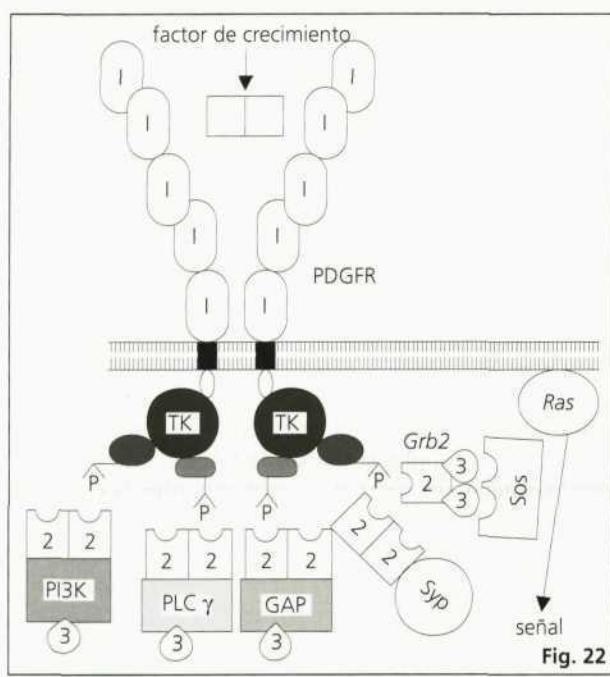
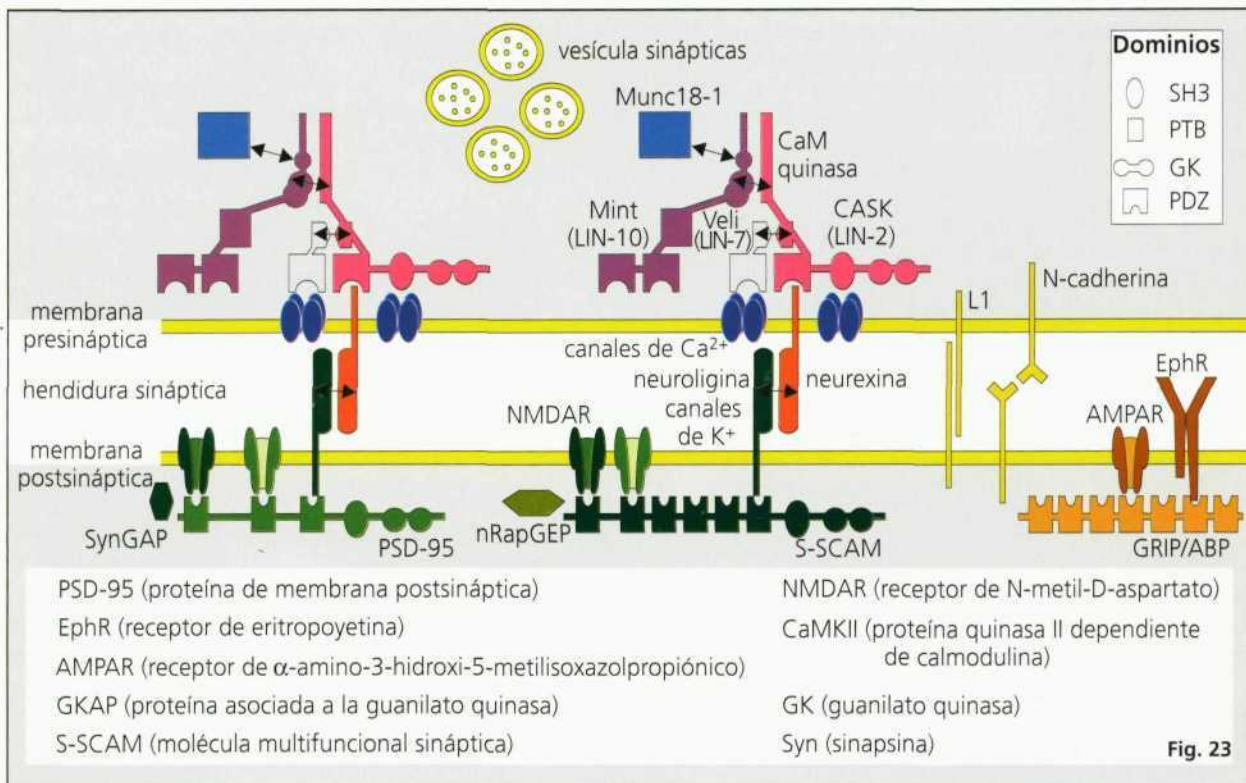


Fig. 21



La figura 22 muestra otro interesante ejemplo a propósito de los mecanismos de *transducción de señales*. Se trata de la recepción del *factor de crecimiento* por su receptor (PDGFR), que posee dominios de tirosina quinasa que cataliza la fosforilación de varios mediadores (*PI 3-quinasa*, *fosfolipasa Cγ* y la *proteína activante de GTPasa*) poseedores de módulos SH2 y SH3.

La figura 23 muestra la actuación de diferentes dominios de proteínas en un ejemplo de interacciones sinápticas. Las *sinapsis* son uniones asimétricas entre neuronas, que transfieren señales desde una neurona presináptica a una célula postsináptica. En las uniones sinápticas, las membranas presinápticas y postsinápticas se conectan por medio de distintas estructuras capaces de funciones especializadas. En las uniones sinápticas se ha descrito una familia de proteínas de membrana: PSD-95/SAP90, PSD-93/capsyn, SAP102 y SAP97/hDLG. Estas proteínas son semejantes a las de otras uniones intercelulares del estilo de ZO-1/-2 y dlg-A. Y todas ellas están compuestas de tres dominios PDZ NH<sub>2</sub>-terminales, un único dominio interior SH3 y un dominio COOH-terminal de guanilato ciclase enzimáticamente inactivo. Los tres dominios PDZ de PSD-95 y proteínas relacionadas interactúan de manera específica: con las secuencias COOH-terminales de los canales de K<sup>+</sup>, con los receptores NMDA2 y con las neuroliginas. En virtud de estas interacciones se consiguen agrupamientos de los canales de K<sup>+</sup> y de receptores NMDA en las sinapsis. Las *neuroliginas* son proteínas de la superficie celular de las neuronas, y están compuestas de un dominio extracelular que recuerda a la *acetilcolinesterasa*, un dominio transmembranar y una secuencia intracelular muy conservada. El dominio extracelular de las neuroliginas interacciona fuertemente con el dominio extracelular de las proteínas de superficie β-neurexinas. Incluso, *neuroliginas* y *β-neurexinas*, localizadas en células independientes, se unen entre sí y dan lugar a uniones intercelulares heterotípicas. Asimismo se ha descrito una



proteína multiligando, S-SCAM (*synaptic scaffolding molecule*), capaz de reunir receptores y proteínas de adhesión celular en las uniones sinápticas.

### PLEGAMIENTO DE LAS PROTEÍNAS

Ya se ha visto cómo las proteínas son cadenas lineales de aminoácidos que adoptan estructuras tridimensionales únicas, lo que les permite llevar a cabo funciones biológicas singulares, y que, dadas las condiciones apropiadas, la mayoría de ellas se pliegan espontáneamente hacia sus estados nativos. Estas cadenas polipeptídicas que forman las proteínas poseen miles de átomos y, por tanto, la posibilidad de millones de interacciones interatómicas. De tal complejidad hubiera parecido imposible realizar predicción alguna acerca de la estructura de las proteínas y de sus mecanismos de plegamiento. Sin embargo, toda la información que una proteína necesita para adquirir su estructura tridimensional se encuentra en la secuencia de aminoácidos.

Así pues, el problema del plegamiento de las proteínas se plantea de manera muy simple: ¿cómo la estructura primaria de las proteínas es capaz de especificar los niveles estructurales tridimensionales superiores? El problema físico tiene un gran interés científico intrínseco, ya que el paso de unas moléculas con un elevadísimo número de grados de libertad a una única estructura tridimensional portadora de funciones singulares es uno de los casos más sencillos de autoorganización biológica. Por otro lado, el problema es de gran importancia práctica en la era genómica, puesto que la interpretación de la enorme información resultante de la secuenciación del DNA ha de requerir la determinación de las estructuras y funciones de las cadenas polipeptídicas codificadas. Y, por tanto, la predicción de las estructuras proteicas será de importancia fundamental en este proceso.

Dado que el número de conformaciones accesibles a una cadena polipeptídica crece exponencialmente con la longitud de la cadena, el punto de partida para la elaboración de un modelo que intente describir el plegamiento de una proteína serán los datos experimentales obtenidos con cadenas cortas, inferiores a los 100 residuos. Y, así, estas reacciones de plegamiento pueden, en general, modelizarse como una transición entre dos estados: un estado desordenado desnaturalizado y otro nativo ordenado. Sin embargo, la cinética de plegamiento de las proteínas de gran longitud parece estar dominado por el abandono de las conformaciones no-nativas de baja energía libre. Este plegamiento de las largas proteínas viene ocasionalmente facilitado por las *chaperonas moleculares*, que impiden la agregación imprópria de las proteínas. Para pasar de un estado no-nativo, sin plegamientos, a un estado nativo de baja energía libre, la proteína ha de atravesar estados de transición de elevada energía libre, cuya heterogeneidad ha sido objeto de repetidos trabajos. En el estado inicial, sin plegamientos, la proteína puede adquirir una de las

muchas conformaciones, mientras que en el estado nativo posee solamente una o pocas conformaciones distintas.

Las medidas experimentales conducentes a la modelización de la reacción de plegamiento de las cadenas polipeptídicas cortas son: la velocidad de plegamiento; la distribución de las estructuras en el estado de transición, deducido de los efectos de las mutaciones sobre la velocidad de plegamiento, y la estructura del estado nativo. Varias líneas de investigación indican que la topología de una proteína determina el mecanismo y la velocidad del plegamiento de las cadenas en mayor medida que las interacciones interatómicas. En primer lugar, los cambios de las largas cadenas en la secuencia de aminoácidos, ya sean de carácter evolutivo o experimental, que no alteran la topología global de una proteína ejercen, por lo general, un efecto muy pequeño sobre la velocidad de plegamiento; lo que sugiere que la evolución no ha optimizado las secuencias de aminoácidos para lograr un plegamiento más rápido. En segundo término, se ha comprobado que las estructuras de los estados de transición son muy insensibles a los cambios profundos de la secuencia, al estudiar los estados de transición de las proteínas con similares estructuras pero diferentes secuencias. Y, en tercer término, las velocidades de plegamiento de las pequeñas proteínas se correlacionan con la topología del estado nativo, y en particular con la separación secuencial media entre los residuos que contactan en la estructura tridimensional, el llamado *orden de contacto*. Así, las proteínas con una buena parte de los contactos entre los residuos cercanos en la secuencia –orden de contacto «bajo»– tienden a plegarse más rápidamente que las proteínas con mayor número de contactos no-locales –orden de contacto «elevado»–. Esta correlación da cuenta de la gran variabilidad de la velocidad de plegamiento, lo que es importante dadas las grandes diferencias en las secuencias y estructuras de las proteínas comparadas. Es decir, las sencillas consideraciones geométricas explican en gran medida las diferencias de las velocidades de plegamiento de las diversas proteínas.

Por otro lado, el importante papel de la *topología del estado nativo* se subraya al considerar el elevado coste entrópico de la formación de contactos entre residuos distantes en la secuencia al comienzo de los plegamientos, debido a que ello restringe el número de conformaciones disponibles a los segmentos implicados. Por esta razón, las interacciones entre los residuos cercanos en la secuencia se encuentran más favorecidas en los plegamientos iniciales que aquellas entre residuos con una separación notable, de forma que, para una topología dada, las interacciones locales tienen lugar más rápidamente que las no-locales. Y, en lógica consecuencia, aquellas topologías sencillas dotadas mayoritariamente de interacciones locales se forman más rápidamente que aquellas con numerosas interacciones no-locales. Finalmente, la importante participación de la topología se ha concluido así mismo de los estudios de los *modelos computacionales de plegamiento de las proteínas*.

Como quiera que las secuencias de las cadenas polipeptídicas determinan las estructuras tridimensionales, resulta que el ordenamiento de los aminoácidos será capaz de determinar tanto la *estabilidad de la molécula* como los *mecanismos de plegamiento*. Sin embargo, mientras que a la estabilidad contribuyen otros factores –como las interacciones interatómicas–, los mecanismos de plegamiento dependen fundamentalmente de las propiedades geométricas del estado nativo.

A partir de estas ideas se han desarrollado dos tipos principales de modelos para la predicción de los mecanismos de plegamiento: la *predicción a partir de la topología* y la *predicción «ab initio»*. En el segundo caso, la predicción consiste en generar el mayor número posible de conformaciones diferentes y calcular la energía de cada una de ellas; y se juzga que la estructura previsible es la conformación de energía más baja. Y para llegar a las estructuras tridimensionales desde las secuencias, hay que utilizar las interacciones entre los átomos en el esqueleto de la molécula, las cadenas laterales y el agua.

#### ERRORES DE PLEGAMIENTO Y ENFERMEDAD

El elevado número de diferentes tipos de proteínas existentes en la naturaleza y, en particular, en la especie humana –alrededor de las 100 000–, su biogénesis a expensas de la información genómica (figura 1) y la extraordinaria variedad de sus funciones (tabla IV), capaces de controlar cada uno de los acontecimientos biológicos, unido a la multiplicidad de bellas formas que presentan en su estado nativo, a los complejos mecanismos de su plegamiento natural y a los procedimientos evolutivos de los sistemas biológicos que aseguran su corrección –o, en caso

contrario, su destrucción, para evitar el daño al organismo huésped– dan cuenta de las numerosas posibilidades de errores de plegamiento de las proteínas y, por tanto, de sus consecuencias y de las posibilidades de disfunción de la maquinaria celular.

A modo de ejemplo, la *fibrosis cística* es un caso en el que una mutación génica ocasiona el plegamiento incorrecto de la proteína responsable del transporte y, de aquí, la imposibilidad de su secreción en la cantidad requerida para el correcto ejercicio de su función. Una situación semejante es la que se presenta en el *enfisema familiar*, debido a que la mutación ocasiona un tráfico incorrecto de las proteínas implicadas. Sin embargo, el mayor interés lo presenta actualmente el conjunto de enfermedades caracterizadas por la conversión de proteínas solubles, o de sus fragmentos, en placas o fibrillas insolubles que se acumulan en una variedad de órganos: hígado, bazo y cerebro. Son las enfermedades que se agrupan como *amiloidosis*; entre ellas, las enfermedades de *Parkinson* y *Alzheimer*, las encefalopatías del tipo *Creutzfeldt-Jakob*, la *diabetes tipo II* y una corta serie de condiciones diversas entre las que se incluye el *insomnio familiar fatal*. Cada enfermedad está asociada a la acumulación de una proteína particular; algunas bien conocidas, como *priones*, *lisozima* y *transferrina* –homotetrámero de 15.87 kDa, elipsoide, presente en la sangre y ligante de hormonas tiroideas–, y también a una cantidad determinada, muy elevada en algunas de las manifestaciones patológicas.

Las fibrillas o placas *amiloideas* son agregados proteínicos que presentan una birrefringencia verde característica bajo luz polarizada y un desplazamiento al rojo en la absorción del rojo Congo; fenómenos ambos atribuidos a la interacción del colorante con las cadenas de la proteína regularmente espaciadas (figura 24). La capacidad de

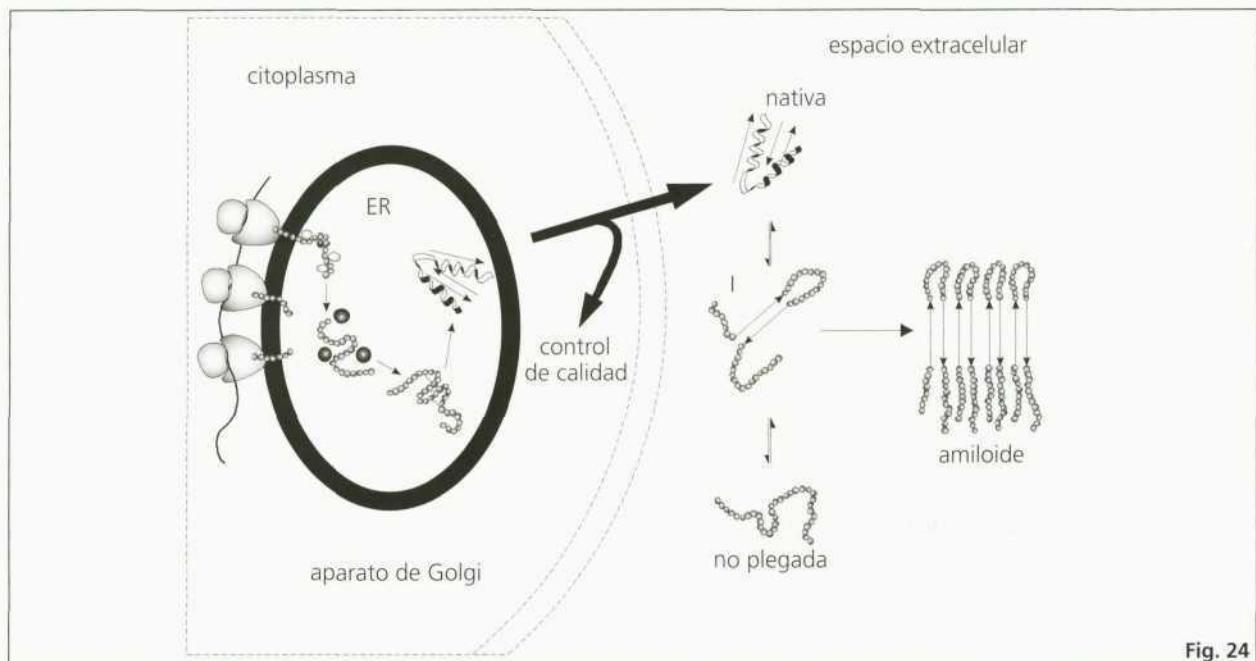


Fig. 24

formar *amiloides* es una propiedad genérica de las cadenas polipeptídicas, interpretada por el hecho de que los enlaces intermoleculares que estabilizan este material están implicados en el esqueleto peptídico común a todas las proteínas. Sin embargo, la formación de fibrillas está restringida a un corto número de proteínas. Y a este propósito queda abierta la cuestión: de un lado, acerca del motivo o de los motivos por los que la mayor parte de las proteínas tienen impedido el acceso en el organismo a este tipo de materiales; de otro, acerca de su formación, favorecida por la mutación de la secuencia de aminoácidos en las proteínas relacionadas con la enfermedad, bajo infecciones, o por el envejecimiento.

Una modificación estructural semejante guarda relación con el conjunto de problemas subyacentes a la *encefalopatía espongiforme bovina*, llamada vulgarmente *enfermedad de las vacas locas*, y su propagación al hombre por vía alimentaria bajo una nueva forma de la *enfermedad de Creutzfeldt-Jakob*, de la cual, al finalizar el año 2000, se habían detectado 79 casos en Gran Bretaña y dos en Francia. Quedan por aclarar las formas de transmisión y los períodos de incubación en el hombre, así como las funciones fisiológicas que las proteínas de los priones sobre la superficie de las neuronas cumplen en la transmisión de ciertas señales celulares, que se alteran profundamente tras la modificación estructural que aparece en la figura 25.

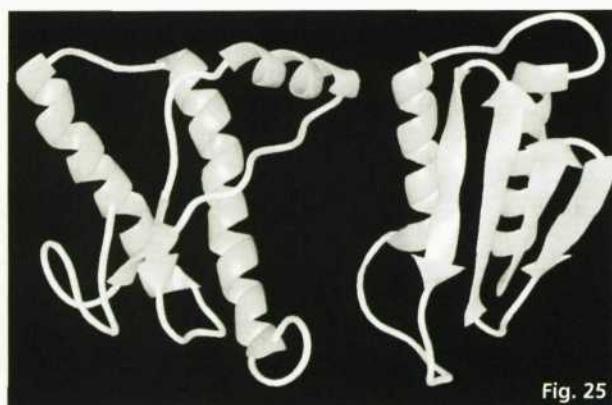


Fig. 25

Estudios llevados a cabo con *Saccharomyces cerevisiae* han conducido al descubrimiento de proteínas con propiedades infectivas, relacionadas con las de los priones de mamíferos. Estos *priones de levaduras*, especie de material genético no cromosómico, son capaces de convertirse *in vitro* en agregados ordenados con todas las características de las fibrillas amiloideas; capacidad de pasar a una forma amiloidea que es transmisible a través de la división celular. Estas propiedades de los *priones de levaduras* y de hongos sugieren que la transferencia de información genética es posible sin la presencia de los ácidos nucleicos. Por otro lado, el mejor conocimiento de las fibrillas amiloideas, y de su formación, redundará en la comprensión de las condiciones patológicas que conducen a muchas de las enfermedades, consecuencia del notorio envejecimiento de

la población moderna, así como en la mejora de las oportunidades para su prevención y tratamiento.

### LA FUNCIÓN DE LAS PROTEÍNAS EN LA ERA POSTGENÓMICA

La disponibilidad en el año 2000 de alrededor de dos docenas de genomas secuenciados y de los perfiles de expresión ha facilitado el desarrollo de métodos computacionales y de nuevas estrategias para encontrar las conexiones funcionales entre las proteínas. Recientemente se han determinado las relaciones funcionales existentes entre las 6 217 proteínas de la levadura *Saccharomyces cerevisiae*, agrupadas a través de las correlaciones de los *perfiles filogenéticos*, los *esquemas de expresión del mRNA* y la *fusión de dominios*. En el Instituto de Biología Molecular de UCLA se ha desarrollado un algoritmo que, mediante estos procedimientos, ha descubierto más de 93 000 conexiones entre proteínas de levadura relacionadas funcionalmente. Las conexiones entre proteínas caracterizadas y no caracterizadas han permitido asignar una función a más de la mitad de las 2 557 proteínas sin caracterización previa. Este tipo de conexiones funcionales se ha establecido entre una familia de proteínas de función desconocida previamente, una proteína cuyas homólogas humanas están implicadas en el cáncer de colon y el prión de levadura *Sup35*.

La figura 26 muestra las estrategias utilizadas para el establecimiento de las conexiones mencionadas en *Saccharomyces cerevisiae*, a la vez que el número de ellas establecidas por cada uno de los procedimientos utilizados. Las *interacciones experimentales* (a) utilizan los datos de la literatura que emplean técnicas como la co-inmunoprecipitación y los procedentes de las bases de datos sobre las interacciones proteína-proteína. La *función metabólica* (b) conecta aquellas proteínas que operan secuencialmente en procesos metabólicos. Los *perfiles filogenéticos* (c) se construyen para conectar las proteínas que han evolucionado de manera correlacionada en los organismos con genoma conocido (figura 15). El método de la *piedra Rosetta*

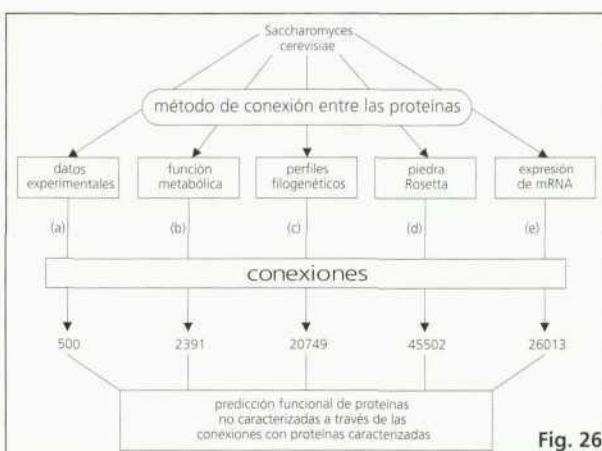


Fig. 26

*setta* (d) conecta las proteínas homólogas que se fusionan en un único gen en otro organismo. La *expresión del mRNA* (e) sirve para correlacionar las proteínas con los niveles de mRNA. Este conjunto de análisis produce redes de conexiones funcionales entre las proteínas y –aunque sea reiterativo insistir en ello– ha alterado fundamentalmente el significado de la idea «función de una proteína».

Otro ejemplo interesante es el que ofrecen los resultados de las conexiones funcionales del *príon de levadura Sup35*. En la figura 27 se ilustra la complicada red de co-

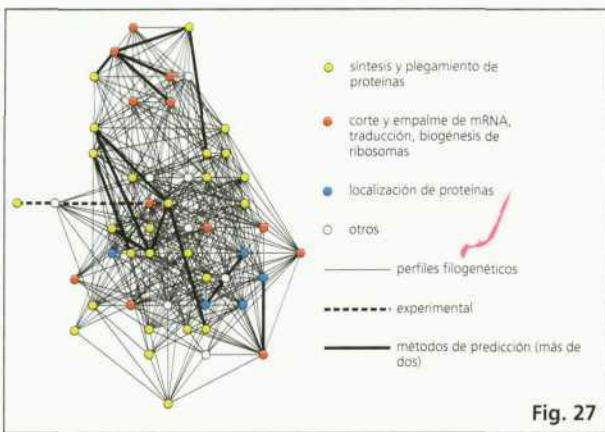


Fig. 27

nexiones de distinto tipo de la proteína del príon de levadura, que actúa en su estado precursor como un factor de liberación de las cadenas traducidas; conexiones con diversas proteínas, como las implicadas en la síntesis y plegamientos de las proteínas –consistente con la función primaria de *Sup35* que interacciona con el ribosoma para liberar la cadena peptídica recién sintetizada–, las proteínas que guían las proteínas nacientes a su destino celular final y con el sistema de chaperoninas (CCT) que colabora en el plegamiento de la actina y los microtúbulos recién sintetizados. En esta red, la mayoría de las conexiones están reconocidas mediante los *perfils filogenéticos*, el *método de la piedra Rosetta* y los *esquemas de expresión del mRNA*. Resulta interesante hacer notar como hecho central de estas redes que la mayor parte de las proteínas interactúan con otras proteínas; y, más particularmente, que las conexiones de la proteína *Sup35*, cuya función es la regulación de la traducción, se establecen con otras relacionadas con la síntesis, el plegamiento y la localización de este tipo de moléculas.

Así pues, estos nuevos métodos para el establecimiento de conexiones funcionales entre las proteínas, a los que, indudablemente, contribuyen el conocimiento progresivo de secuencias genómicas completas y de los datos sobre la expresión de mRNAs, originan redes de interacciones extraordinariamente complejas.

Como resumen de esta visión de la función de las proteínas en la era postgenómica hay que hacer notar que, clásicamente, la función de las proteínas estaba basada en la

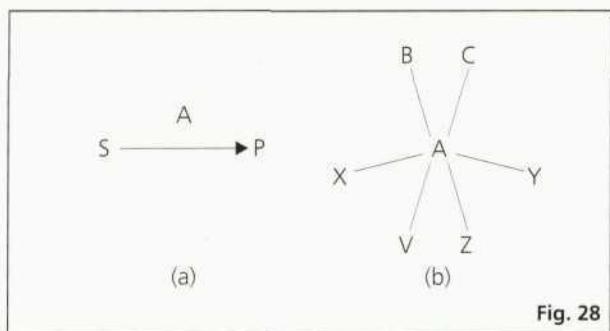


Fig. 28

acción individual de cada una de ellas, como puede ser la catálisis de una reacción determinada, o la unión de ligandos específicos de mayor o menor tamaño (figura 28a). Frente a esta idea «restringida» de función –frecuentemente conocida como *función molecular*–, se considera en la actualidad la idea amplia de la función según la cual *una proteína se define como un elemento en el conjunto de sus interacciones* (figura 28b). Esta idea amplia de la función de las proteínas se conoce como *función contextual* o *función celular*, que intenta significar cómo, *en el seno de la materia viva, cada proteína funciona como elemento de una amplia red de interacciones moleculares*. Bajo esta idea será posible comprender las funciones celulares de aquellas proteínas no caracterizadas por medio de sus interacciones con proteínas caracterizadas. Y así, las redes de conexiones ofrecen una nueva visión del significado de la *función*, y una comprensión más profunda del funcionamiento de las células.

#### LA FUNCIÓN DE LAS PROTEÍNAS Y LAS SECUENCIAS GENÓMICAS

Uno de los nuevos métodos para el estudio de las interacciones proteína-proteína utiliza procedimientos computacionales para el análisis de las secuencias genómicas sobre la base de que algunos pares de proteínas conexionadas guardan homología con las de otro organismo, fusionadas en una única cadena polipeptídica. En el departamento de Bioquímica de UCLA se han descrito varios métodos para este *análisis de la fusión de dominios*. Algunos de ellos utilizan compilaciones de pares de proteínas descritas en la literatura; otro método computacional es el de los *perfils filogenéticos* (figura 15), ya mencionado, que detecta las interacciones funcionales analizando la evolución correlacionada de las proteínas. La investigación de las secuencias de algunos genomas ha puesto de manifiesto 6 809 interacciones proteína-proteína en *Escherichia coli* –a partir de las 4 290 secuencias de proteínas de su genoma– y 45 502 interacciones en la levadura. Y muchos de estos pares de conexiones están funcionalmente relacionados. En la figura 29 aparecen algunos de los pares de proteínas que interactúan en *E. coli* y que se encuentran fusionadas en una única cadena de otros organismos.

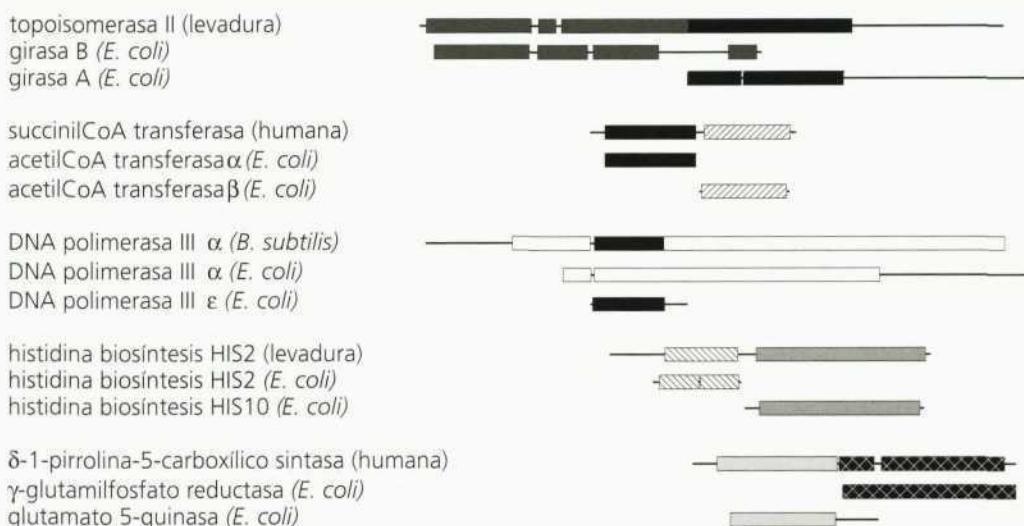


Fig. 29

El análisis y la reconstrucción de procesos metabólicos en *E. coli* por los métodos descritos de *fusión de dominios* ha permitido el descubrimiento de *interacciones acopladas*. En la figura 30 aparecen los esquemas de las transformaciones metabólicas correspondientes a las síntesis

del ácido shiquímico (A) y de las purinas (B), ordenadas según el método tradicional de la acción de las sucesivas enzimas sobre los correspondientes metabolitos; algunas de estas enzimas están conexionadas con otras del mismo proceso y aparecen en negrita. Los esquemas C y D re-

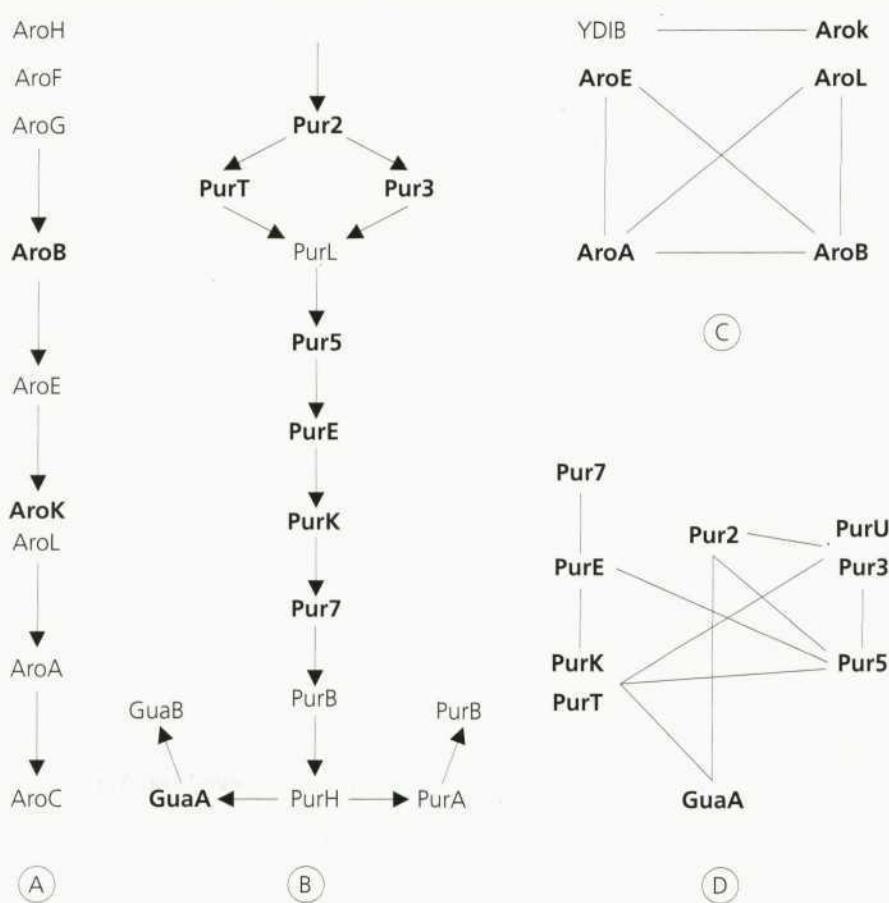


Fig. 30

presentan las transformaciones sugeridas por las interacciones habidas, a la vez, entre enzimas secuenciales del proceso y entre miembros más distantes, posiblemente a través de complejos multienzimáticos.

Así pues, la información genómica ha abierto nuevas vías a descubrimientos bioquímicos acerca de nuevos procesos o de nuevos complejos de proteínas en los diferentes organismos.

### INGENIERÍA DE PROTEÍNAS

Durante el tercer cuarto del siglo XX tuvo lugar la primera gran etapa del definitivo conocimiento estructural de las proteínas, y su aplicación a una amplia colección de proteínas recién descubiertas en los campos de la farmacología, la agricultura y la industria –por ejemplo, la insulina, las toxinas de insectos y las enzimas industriales-. Hubo que llegar, sin embargo, al último cuarto de siglo para alcanzar el descubrimiento de la tecnología del DNA-recombinante y, con ella, la posibilidad de producción di-

recta artificial de proteínas humanas, comienzo de los ulteriores desarrollos de la *ingeniería de proteínas*. A ellos han contribuido toda la serie de procedimientos encaminados al mejor conocimiento de la estructura de las proteínas y al de las relaciones *estructura-función*. Entre otros, la averiguación de la secuencia de aminoácidos, la creación y modificación de grupos funcionales, y la determinación de las estructuras tridimensionales por difracción de rayos X y la espectroscopía de resonancia magnética, con las que se ha podido averiguar la topografía de las superficies complementarias entre las que se establecen las interacciones ligando-proteína y proteína-proteína, determinantes de gran parte de las funciones biológicas. Interacciones que pueden examinarse mediante la visión gráfica computarizada que permite modular parámetros del tipo de especificidad, estabilidad y afinidad de unión, como consecuencia de hipotéticas modificaciones estructurales, incluidos los cambios mutacionales de aminoácidos.

Fruto de todo ello ha sido la aplicación de la modificación estructural de proteínas –la *ingeniería de proteínas*– al diseño de variaciones en la función de *enzimas, anticuer-*

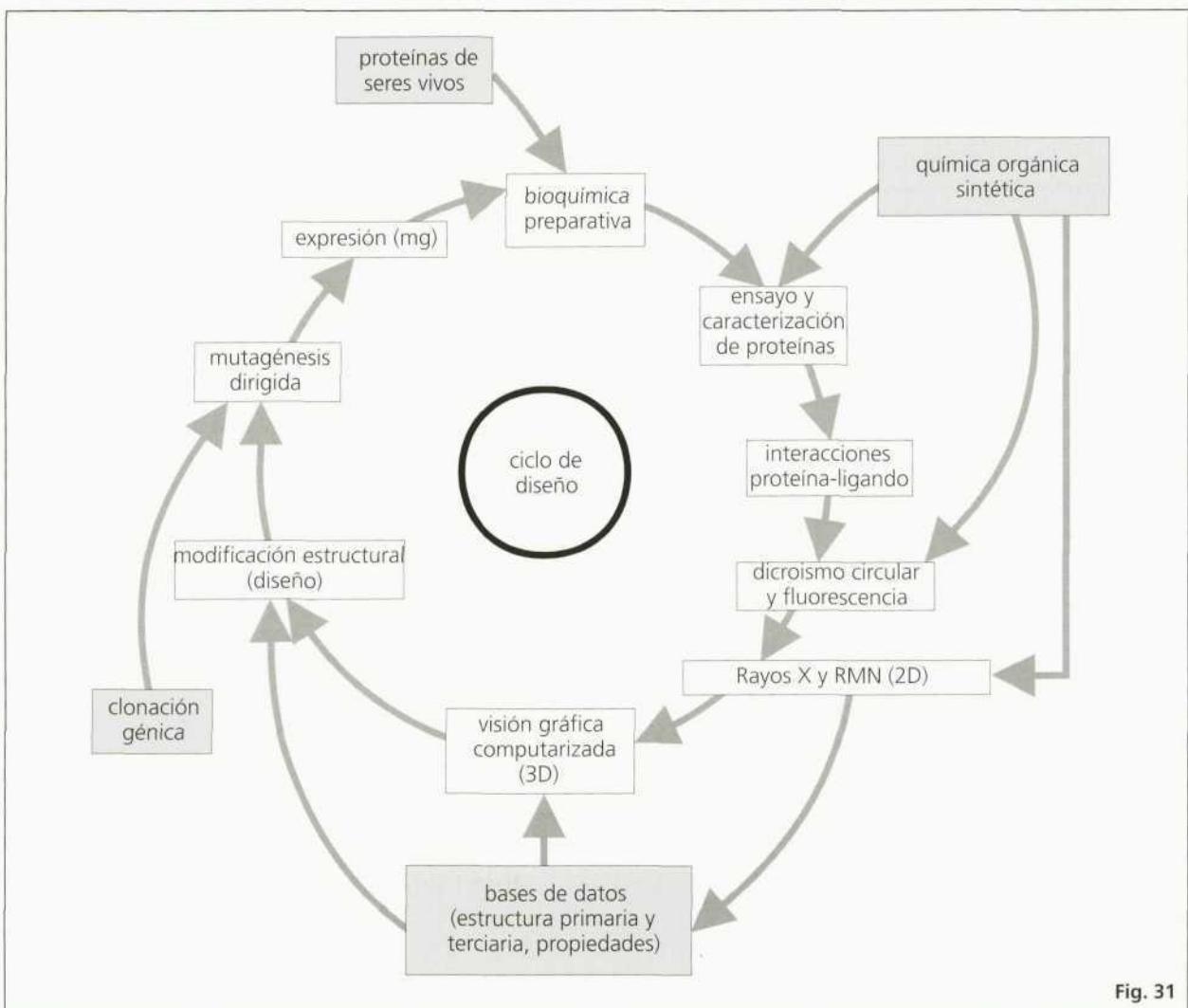


Fig. 31



Fig. 32

pos, ligandos, receptores de membrana, proteínas transductoras de señales, etc., y a las posibilidades de modificación. Todo este conjunto de aproximaciones al estudio de las relaciones estructura-función constituye el llamado *ciclo de diseño* (figura 31), fundamental para el tratamiento interdisciplinar racional de las proteínas, y conducente a sus aplicaciones específicas.

Dentro de los esquemas de diseño de nuevas actividades de proteínas, la creación de nuevas *actividades de unión de ligandos* tiene una significación especial, ya que otorga a estas moléculas la funcionalidad propia de todos los procesos de reconocimiento molecular, al estilo de los implicados en los numerosos mecanismos de transducción de señales, y de sus variadas aplicaciones comerciales como reactivos de diagnóstico o agentes terapéuticos de diversa índole –nuevos anticuerpos, pequeñas proteínas útiles como ligandos de receptores de superficie, etc.–. En esta dirección se han descrito en los últimos años diversas técnicas para generar péptidos y proteínas con especiales actividades de unión de ligandos. Entre estas técnicas, los *fagos filamentosos* reúnen características singulares para la selección de repertorios de estos péptidos o proteínas que, fusionados con las proteínas de la cubierta, pueden expresarse bajo una conformación funcional. A este objeto, los genes codificantes, cuyas secuencias se conocen o determinan, se encapsulan dentro de la cubierta del fago, y los péptidos o proteínas se aislan y purifican por afinidad sobre un lecho inmovilizado. De esta manera, se han originado muchos archivos de péptidos (*peptidotecas*) de diferentes tamaños (figura 32a) –en los que la variación se introduce en un único fragmento–, que pueden utilizarse como ligandos estructurales o para modelizar las actividades biológicas de moléculas de mayor magnitud. El primero de estos archivos de fagos estuvo constituido por péptidos randomizados de seis residuos, fusionados a la posición N-terminal de la proteína de adsorción del fago pIII. Posteriormente se han generado otros muchos

archivos de péptidos de diferente tamaño –de hasta 40 residuos– y conformación.

De igual manera, por medio de los fagos pueden crearse los archivos de proteínas plegadas, cuyos residuos se randomizan simultáneamente en varias regiones discontinuas de la cadena polipeptídica. Y estos residuos randomizados, en el seno de la proteína plegada, forman lo que semeja un parche continuo sobre la superficie de la proteína. Uno de los ejemplos más representativos de esta técnica (*patch engineering*) es el archivo de anticuerpos (figura 32b), en el que la variación se logra por randomización de seis diferentes regiones de la cadena polipeptídica. Estas regiones, a través de plegamientos, se ponen en contacto sobre una cara de la proteína para originar un repertorio de fragmentos Fab o de fragmentos scFv (*single chain Fv*).

Esta técnica ha sido utilizada recientemente para la creación de sitios adicionales de unión en varias proteínas, por ejemplo, en un *inhibidor de α-amilasa*, de 74 residuos, procedente de *Streptomyces tendae*, y en *citocromo β<sub>562</sub>*, de 106 residuos, que se pliega en una estructura en tirabuzón de cuatro hélices, estabilizada por unión del hemo (figuras 11 y 12). Y así, dos lazos superficiales, constituidos por los residuos 37-41 y 88-92, que no participan en los contactos con el hemo, forman un parche de 100 Å<sup>2</sup> (figura 33) que ha podido ser randomizado en otra proteína con actividad de unión de ligandos modificada o adicional. Estos repertorios pueden ponerse de manifiesto en fagos, y, mediante selección, aislar proteínas con nuevas actividades de unión.

#### IDENTIFICACIÓN Y ANÁLISIS DE PROTEÍNAS

Si la *proteómica* se ha asociado a la exhibición de un gran número de proteínas a partir de un organismo o una línea celular determinados, utilizando los sistemas de se-



paración más eficientes, tal como la *electroforesis bidimensional en gel de poliacrilamida*, habrá que señalar que ya en la década de los setenta se comenzaron a construir bases de datos de las proteínas a partir de las separaciones logradas mediante este tipo de técnicas. Sin embargo, no era fácil en esa fecha la rápida caracterización analítica de las proteínas, y hubo que esperar, en la década de los noventa, al desarrollo de un poderoso método analítico, la *espectrometría de masas*, con el que solucionar todas las limitaciones anteriores. Este desarrollo, unido a la disponibilidad pública de las secuencias genómicas, ha desembocado en el análisis funcional de los productos de los genes; área que, fundamentalmente, cubre el nuevo campo de la *proteómica*.

#### BIBLIOGRAFÍA

- ABBOTT, A., «How to spot a protein in a crowd», *Nature* 402, 1999, págs. 716-720.
- ALM, E. y BAKER, D., «Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures», *Proc. Natl. Acad. Sci. USA* 96, 1999, págs. 11305-11310.
- ANDERSON, N. G. y ANDERSON, N. I., «Twenty years of two-dimensional electrophoresis: past, present and future», *Electrophoresis* 17, 1996, págs. 443-453.
- BAKER, D., «A surprising simplicity to protein folding» *Nature* 405, 2000, págs. 39-42.
- BLUNDELL, T. L., «Problems and solutions in protein engineering-towards rational design», *TIBTECH* 12, 1994, págs. 145-148.
- BORMAN, S., «Biochemical Genomics», *C&EN* 5, 1999.
- CHOTHIA, C., «One thousand families for the molecular biologist», *Nature* 357, 1992, págs. 543-544.
- EISENBERG, D., MARCOTTE, E. M., XENARIOS, I. y YEATES, T. O., «Protein function in the post-genomic era», *Nature* 405, 2000, págs. 823-826.
- GOODSELL, D. S. y OLSON, A. J., «Soluble proteins: size, shape and function», *TIBS* 18, 1993, págs. 65-68.
- HIRAO, K., HATA, Y., IDE, N., TAKEUCHI, M., IRIE, M., YAO, I., DEGUCHI, M., TOYODA, A., SUDHOFF, T. C. y TAKAI, Y., «A Novel Multiple PDZ Domain-containing Molecule Interacting with N-Methyl-D-aspartate Receptors and Neuronal Cell Adhesion Proteins», *J. Biol. Chem.* 273, 1998, págs. 21105-21110.
- HOLM, L. y SANDER, C., «Mapping the Protein Universe», *Science* 273, 1996, pág. 595.
- IRIE, M., HATA, Y., TAKEUCHI, M., ICHTCHENKO, K., TOYODA, A., HIRAO, K., TAKAI, Y., ROSAHL T. W. y SÜDHOF, T. C., «Binding of neuroligins to PSD-95», *Science* 277, 1997, págs. 1511-1515.
- LANDSCHULZ, W. H., JOHNSON, P. F. y MCKNIGHT, S. L., «The Leucine Zipper: A Hypothetical Structure Common to a New Class of DNA Binding Proteins», *Science* 240, 1988, págs. 1759-1764.
- MARCOTTE, E. M., PELLEGRINI, M., HO-LEUNG NG, RICE, D. W., YEATES, T. O. y EISENBERG, D., «Detecting Protein Function and Protein-Protein Interactions from Genome Sequences», *Science* 285, 1999, págs. 751-753.
- —, PELLEGRINI, M., THOMPSON, M. J., YEATES, T. O. y EISENBERG, D., «A combined algorithm for genome-wide prediction of protein function», *Nature* 402, 1999, págs. 83-86.
- MARTÍN MUNICIO, A., «Medicamentos viejos para enfermedades nuevas», *Horizontes culturales. Las fronteras de la ciencia*, Real Academia de Ciencias, Madrid, 1998.
- —, «Presente y futuro de la biotecnología», *Horizontes culturales. Las fronteras de la ciencia*, Real Academia de Ciencias, Madrid, 1999.
- O'FARRELL, P. H., «High resolution two-dimensional electrophoresis of proteins», *J. Biol. Chem.* 250, 1975, págs. 1007-1021.
- PANDEY, A. y MANN, M., «Proteomics to study genes and genomes», *Nature* 405, 2000, págs. 837-846.
- PELLEGRINI, M., MARCOTTE, E. M., THOMPSON, M. J., EISENBERG, D. y YEATES, T. O., «Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles», *Proc. Natl. Acad. Sci. USA* 96, 1999, págs. 4285-4288.

- SALI, A., «Functional links between proteins», *Nature* 402, 1999, págs. 23-26.
- SCHEIFFELE, P., FAN, J., CHOIH, J., FETTER R. y SE-RAFINI, T., «Neuroligin Expressed in Nonneuronal Cells Triggers Presynaptic Development in Contacting Axons», *Cell* 101, 2000, págs. 657-669.
- SCHOMBURG, D., «A broad perspective on protein engineering and complementary technologies», *TIBTECH* 13, 1995, págs. 45-47.
- SMITH, G., «Patch engineering: a general approach for creating proteins that have new binding activities», *TIBS* 23, 1998, págs. 457-460.
- SONG, J. Y., ICHTCHENKO, K., SÜDHOF, T. C. y BRO-SE, N., «Neuroligin 1 is a postsynaptic cell-adhesion molecule of excitatory synapses», *Proc. Natl. Acad. Sci. USA* 96, 1999, págs. 1100-1105.
- THIEFFRY, D. y SARKAR, S., «Forty years under the central dogma», *TIBS* 23, 1998, págs. 312-316.