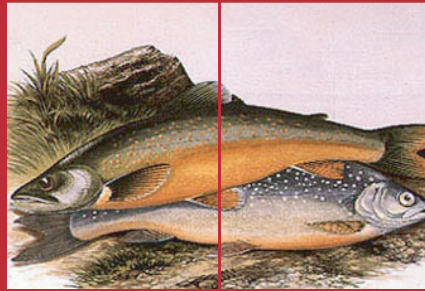


PROTEÍNAS

Estructuras fascinantes

Dr. Javier Santos



Colección: LAS CIENCIAS NATURALES Y LA MATEMÁTICA

Colección: LAS CIENCIAS NATURALES Y LA MATEMÁTICA

PROTEÍNAS

Estructuras fascinantes

Javier Santos

ADVERTENCIA

La habilitación de las direcciones electrónicas y dominios de la web asociados, citados en este libro, debe ser considerada vigente para su acceso, a la fecha de edición de la presente publicación. Los eventuales cambios, en razón de la caducidad, transferencia de dominio, modificaciones y/o alteraciones de contenidos y su uso para otros propósitos, queda fuera de las previsiones de la presente edición -Por lo tanto, las direcciones electrónicas mencionadas en este libro, deben ser descartadas o consideradas, en este contexto-.

Distribución de carácter gratuito.

a u t o r i d a d e s

PRESIDENTE DE LA NACIÓN

Dra. Cristina Fernández de Kirchner

MINISTRO DE EDUCACIÓN

Dr. Alberto E. Sileoni

SECRETARIA DE EDUCACIÓN

Prof. María Inés Abrile de Vollmer

DIRECTORA EJECUTIVA DEL INSTITUTO NACIONAL DE
EDUCACIÓN TECNOLÓGICA

Lic. María Rosa Almandoz

DIRECTOR NACIONAL DEL CENTRO NACIONAL DE
EDUCACIÓN TECNOLÓGICA

Lic. Juan Manuel Kirschenbaum

DIRECTOR NACIONAL DE EDUCACIÓN TÉCNICO PROFESIONAL Y
OCUPACIONAL

Ing. Roberto Díaz

Ministerio de Educación.
Instituto Nacional de Educación Tecnológica.
Saavedra 789. C1229ACE.
Ciudad Autónoma de Buenos Aires.
República Argentina.
2009

PROTEÍNAS

Estructuras fascinantes

Javier Santos



Colección: LAS CIENCIAS NATURALES Y LA MATEMÁTICA

Colección "Las Ciencias Naturales y la Matemática".
Director de la Colección: Juan Manuel Kirschenbaum
Coordinadora general de la Colección: Haydeé Noceti.

Queda hecho el depósito que previene la ley N° 11.723. © Todos los derechos reservados por el Ministerio de Educación, Ciencia y Tecnología - Instituto Nacional de Educación Tecnológica.

La reproducción total o parcial, en forma idéntica o modificada por cualquier medio mecánico o electrónico incluyendo fotocopia, grabación o cualquier sistema de almacenamiento y recuperación de información no autorizada en forma expresa por el editor, viola derechos reservados.

Industria Argentina

ISBN 978-950-00-0707-8

Director de la Colección:
Lic. Juan Manuel Kirschenbaum

**Coordinadora general y académica
de la Colección:**

Prof. Ing. Haydeé Noceti

Diseño didáctico y corrección de estilo:

Lic. María Inés Narvaja
Ing. Alejandra Santos

Coordinación y producción gráfica:
Tomás Ahumada

Diseño gráfico:
Augusto Bastons

Ilustraciones:
Diego Gonzalo Ferreyro

Retoques fotográficos:
Roberto Sobrado

Diseño de tapa:
Tomás Ahumada

Administración:
Cristina Caratozzolo
Néstor Hergenrether

Nuestro agradecimiento al personal
del Centro Nacional de Educación
Tecnológica por su colaboración.

Santos , Javier

Las proteínas / Javier Santos; dirigido por Juan Manuel
Kirschenbaum.

- 1a ed. - Buenos Aires: Ministerio de Educación de la Nación. Instituto
Nacional de Educación Tecnológica, 2009.

192 p. ; 24x19 cm. (Las ciencias naturales y la matemática / Juan Manuel
Kirschenbaum.)

ISBN 978-950-00-0707-8

I. Biología. Educación Secundaria.

I. Kirschenbaum, Juan Manuel, dir.

II. Título

CDD 570.712

Fecha de catalogación: 14/07/2009

Impreso en Anselmo L. Morvillo S. A., Av. Francisco Pienovi 317 (B1868DRG),
Avellaneda, Pcia. de Buenos Aires, Argentina.

Tirada de esta edición: 100.000 ejemplares



Dr. Javier Santos

El Autor

Javier Santos nació el 22 de diciembre de 1972. En su infancia, orientado por su maestro Otto Fenninger se interesó por la Biología y especialmente por las Aves. Estudió Ciencias Biológicas en la Universidad de Buenos Aires entre 1992 y 1998, años en los que conoció al profesor Dr. Daniel Goldstein quien rápidamente le contagió su fascinación por las proteínas. Inició su camino en el estudio de las proteínas en el año 1996 en el Laboratorio de Expresión y Plegamiento de Proteínas dirigido por el Dr. Mario R. Ermácora. Obtuvo el título de Doctor de la Universidad de Buenos Aires en el año 2004.

Actualmente es Profesor Ordinario en el Área de Bioquímica en la Universidad de Quilmes y, desde el año 2005, es Investigador Asistente del CONICET. Se desempeña como investigador en el IQUIFIB (UBA-CO-NICET). Participó en proyectos biotecnológicos y actualmente se dedica al estudio de los mecanismos moleculares que conducen a la consolidación de la estructura tridimensional de las proteínas.

AGRADECIMIENTOS:

Agradezco profundamente a Otto Fenninger, a Daniel Goldstein y a Mario Ermácora. A ellos, a Luís, a Vero y a las personas que han compartido conmigo el trabajo en el laboratorio les dedico este viaje por las fascinantes proteínas.

ÍNDICE

Primera Parte

Estructura Proteica	8
• El problema de la escala	11
• Representación de la estructura proteica (Primera Parte)	11
• Los aminoácidos y sus propiedades	13
• El enlace peptídico	15
• Representación de la estructura proteica	18
• Libertad conformacional, ángulos torsionales	21
• Organización jerárquica y modular	22
• Estructuras helicoidales	23
• Las hélices 3_{10} y las hélices π	27
• Las láminas beta (β)	27
• Estructura super-secundaria y motivos estructurales	28
• Puentes disulfuro y tioles libres	29
• Bases biofísicas de la estabilidad de la estructura proteica	31
• El pH y las proteínas	33
• Interacciones de van der Waals	36
• Campo de fuerzas	39
• Efecto hidrofóbico	40
• Estructura Terciaria, dominios, topología	42
• Familias de proteínas y proteínas multi dominio	44
• Familias y superfamilias (homólogos o análogos)	46
• Estructura cuaternaria, subunidades	46

Segunda Parte

El Proceso de Plegado Proteico	48
• Equilibrio de desplegado $N \leftrightarrow U$	54
• ¿Estado nativo o estado desplegado?	64
• ¿Otros estados? Estados intermedios y parcialmente plegados	65
• Glóbulos fundidos	66
• Desplegado por temperatura	68
• El ΔC_P y el ΔS_{AUN}	70
• Desplegado por pH	71
• La reacción de plegado proteico seguida en función del tiempo	72
• El estado de transición de la reacción de plegado	73
• Paradoja de Levinthal	80
• Modelos de plegado: Teorías para el proceso de plegado	81

• El modelo jerárquico	82
• Rompecabezas	83
• El colapso hidrofóbico	83
• Nucleación condensación	84
• Plasticidad del mecanismo de plegado	84
• Paisajes energéticos y embudos de plegado	85
Tercera Parte	
Producción de proteínas en el laboratorio	88
• El código genético. Genes. Transcripción y traducción	88
• Producción de proteínas recombinantes en bacterias	93
• Tecnología del DNA recombinante	93
• ¿Por qué la proteína de interés puede hallarse en la fracción insoluble?	100
Cuarta Parte	
Estudio de la conformación proteica	106
• Geles de poliacrilamida	106
• Cromatografía de exclusión molecular	109
• Interacción de las proteínas con la luz	117
• El ultravioleta lejano: El enlace peptídico.	120
• El ultravioleta cercano: Los residuos aromáticos.	121
• Fluorescencia	126
• Apagado y apagadores de fluorescencia	131
• Dicroísmo circular	135
• Espectrometría de masa	142
• ¿Qué es un cristal de proteínas?	146
• ¿Qué es la difracción de rayos X?	148
• ¿Qué son los rayos X?	148
Quinta Parte	
Conclusión	154
Guía de problemas	161
Respuestas de problemas	167
Glosario	182
Abreviaturas	191

Estructura Proteica

Primera Parte

Una macromolécula es, concretamente, una molécula grande. Una macromolécula biológica es una molécula grande y compleja que tiene una función biológica. Las proteínas son macromoléculas biológicas. Están formadas por miles de átomos.

Las proteínas tienen una estructura tridimensional (3D) compleja:

Están formadas por una o más cadenas de aminoácidos. Son polímeros de aminoácidos. Los aminoácidos de cada cadena están unidos por enlaces químicos conocidos como enlaces peptídico.

El conocimiento acabado de la estructura covalente de las proteínas llevó años de discusiones y, la más cruda aplicación del método científico basado en la prueba experimental, los modelos y las ideas.

Hoy en día sabemos que la secuencia en la que están dispuestos los aminoácidos de una proteína es crucial, es extremadamente informativa y determina la geometría y, por lo tanto, la estructura 3D de esta proteína. Dicho de otro modo, la secuencia determina el recorrido de la cadena de aminoácidos en el espacio y la topología de la proteína. Sabemos que las secuencias de aminoácidos que tienen la capacidad de plegarse para dar proteínas son sólo una parte pequeña de todas las secuencias de aminoácidos posibles. *Por esto, las proteínas son un subconjunto muy especial dentro del conjunto de los polipéptidos.*

Uno de los problemas más importantes que tratamos de resolver es que, por ahora, tenemos sólo algunos indicios acerca de cómo está codificada esa información que determina la estructura en la secuencia de aminoácidos. También tenemos algunos indicios de que parte de la secuencia es importante para la adquisición de estructura 3D y que otra parte no lo es.

Las proteínas pueden cambiar de conformación, pueden plegarse y desplegarse, pueden desorganizarse. El número de conformaciones posibles que una proteína, con una secuencia dada, puede adoptar es increíblemente alto, debido a que muchas de las uniones entre átomos tienen cierta libertad para rotar. Sin embargo, cada proteína en condiciones adecuadas de temperatura,

Recordar que, por ejemplo, una molécula de agua está formada por 3 átomos, dos de hidrógeno y uno de oxígeno, y una de etanol, alcohol etílico, por 9 átomos dos átomos de carbono 6 de hidrógeno y uno de oxígeno.

Por esto, también, se las llama polipéptidos: polímeros de aminoácidos unidos mediante enlaces peptídicos.

En los años 1930-1940 se daban, simultáneamente, el inicio de la cristalografía de proteínas, el inicio del estudio del proceso del plegado proteico, la aplicación de conceptos de química y física de pequeñas moléculas; todo junto constituyó un increíble avance en la comprensión de las bases de la estructura macromolecular. Resuenan nombres como A. E. Mirsky y Linus Pauling y sus estudios sobre la estructura proteica, el estado nativo (ver más abajo), el estado desplegado y la estabilización por puentes de hidrógeno intramoleculares de la conformación proteica.

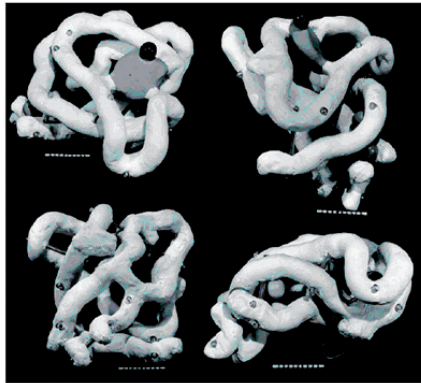
por ejemplo a 25°C, y solvente, el medio en el que está, por ejemplo, agua, adquiere espontáneamente una única conformación muy bien definida. En esas condiciones esa estructura es, energéticamente, la más estable y tiene una organización extrema. A esta estructura la llamamos estructura nativa. Nativa porque es la estructura que, en principio, no sólo está bien plegada, sino que adquiere la proteína tanto en un tubo de ensayo, completamente, aislada de otras moléculas como en el entorno biológico, por ejemplo, el interior de una célula. En el caso de la mayor parte de las proteínas, la función biológica está asociada a la estructura nativa.

Hoy en día podemos decir con certeza que uno de los avances más emocionantes de la ciencia fue la posibilidad de determinar la estructura de una proteína con alta resolución. Contar con estructuras, brinda la posibilidad de entender mecanismos biológicos de enfermedades y diseñar fármacos racionalmente, por indicar dos de las tantas e importantes aplicaciones.

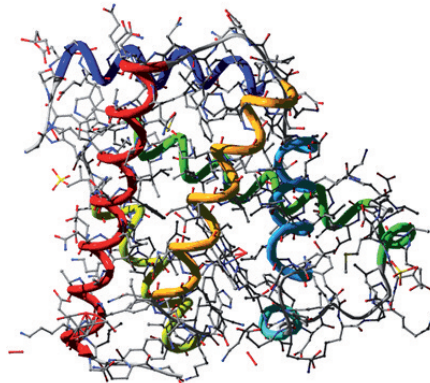
En la actualidad existen únicamente dos métodos que permiten hacerlo. La cristalografía de rayos X y la espectroscopía de resonancia magnética nuclear (RMN). Como todo, cada método tiene ventajas y desventajas. En ambos casos los científicos (cristalógrafos, en un caso y espectroscopistas, en el otro) usan muchas (muchísimas) piezas de información (tal cual un detective) para crear un modelo de la macromolécula a escala atómica. Esas piezas son, efectivamente, *los datos experimentales* acerca de la estructura de la macromolécula. Los datos experimentales son probablemente lo más importante con que cuenta un científico. Para los cristalógrafos estos datos son los patrones de difracción de rayos X que se producen como consecuencia de la dispersión de fotones de alta energía cuando el haz de rayos X impacta con un cristal de las proteínas, más específicamente, cuando los fotones son desviados por la presencia de los electrones de las macromoléculas que forman el cristal. Para los espectroscopistas de RMN son restric-

Los experimentos de Christian B. Anfinsen, en los años 1950-1970 no dejaron dudas sobre este punto. La elaboración de la hipótesis termodinámica, que describe a la conformación nativa como la de menor energía y que detalla al proceso de plegado como un proceso espontáneo, sin duda, marcó el inicio de un largo e increíble camino que tuvo desde su inicio como objeto comprender cómo las proteínas se pliegan en sus conformaciones tridimensionales y cómo la secuencia determina en condiciones adecuadas la adquisición de la estructura final.

Por otra parte, los cálculos propuestos por Cyrus Levinthal permitieron postular que una proteína no encuentra su conformación mediante una búsqueda al azar, ya que como veremos, esto le demandaría tiempos extremadamente largos, no compatibles para nada con los tiempos biológicos. Por el contrario, las proteínas se pliegan rápidamente. Muchas de ellas, incluso, en fracciones de milisegundos (la milésima parte de un segundo) y hasta en microsegundos (la millonésima parte de un segundo). A este planteo, se lo conoce con el nombre de paradoja de Levinthal y, a partir del mismo, es que surge la necesidad de entender cómo se lleva adelante esta búsqueda conformacional, la búsqueda de la conformación de mínima energía. Más adelante también examinaremos cuidadosamente las distintas alternativas que se han propuesto hasta el momento, lo que ha quedado más que claro desde aquel entonces es que la búsqueda de esta conformación no es un proceso aleatorio.



Estructura de la mioglobina publicada por J. Kendrew y colaboradores en el año 1959



Estructura de la mioglobina publicada por H. P. Hersleth y colaboradores en el año 2008

• *Figura 1-1. Comparación de dos modelos de la mioglobina. El modelo de la derecha construido por J. Kendrew y colaboradores en el año 1959. El modelo de la izquierda fue publicado por H.P. Hersleth en año 2008. La imagen de este último fue preparada usando el programa de visualización de estructuras SwissPDBviewer*

ciones de distancias y ángulos entre núcleos atómicos. Una tercera metodología, la microscopía electrónica permite obtener imágenes de la forma general de una macromolécula o de un complejo de macromoléculas. Es usada en conjunto con las anteriormente mencionadas. En este caso los datos son las imágenes obtenidas mediante el estudio del impacto de electrones sobre la macromolécula.

En la actualidad existen bases de datos de estructuras obtenidas por estas técnicas y, su acceso es relativamente sencillo y gratuito. Una de las más importantes es la PDB, o *Protein data Bank* (en español Banco de Datos de Proteínas) un portal en el que podemos encontrar miles de estructuras.

Cada modelo experimental se deposita en forma de un enorme archivo con las coordenadas espaciales de cada uno de los átomos. El crecimiento constante de esta base de datos es consecuencia del trabajo de numerosos científicos. En esta base de datos podemos encontrar estructuras de proteínas y ácidos nucleicos relacionadas con los procesos fundamentales en biología. También podemos encontrar estructuras de complejos macromoleculares (más de una molécula) Nos podemos encontrar, por ejemplo, con la estructura del ribosoma completo o incluso con la estructura de un virus. La visualización de estas estructuras puede hacerse utilizando programas computacionales de distribución gratuita. Vale la pena aprender algunas cosas más mientras lo hacemos.

Los avances técnicos en los últimos 50 años para la obtención de estructuras han sido gigantescos. Esto ha permitido obtener modelos más detallados y con mejor resolución. En la Figura 1-1 se puede ver cómo en las primeras estructuras determinadas por cristalografía de rayos X únicamente podía diferenciarse el trazado general de la cadena polipeptídica en el espacio. Con las estructuras cristalográficas más recientes podemos localizar cada uno de los átomos de la cadena polipeptídica e incluso moléculas de agua, e iones asociados a la estructura. En casos de extrema resolución también se pueden ubicar los átomos de hidrógeno.

El problema de la escala: Distancias e intervalos de tiempo

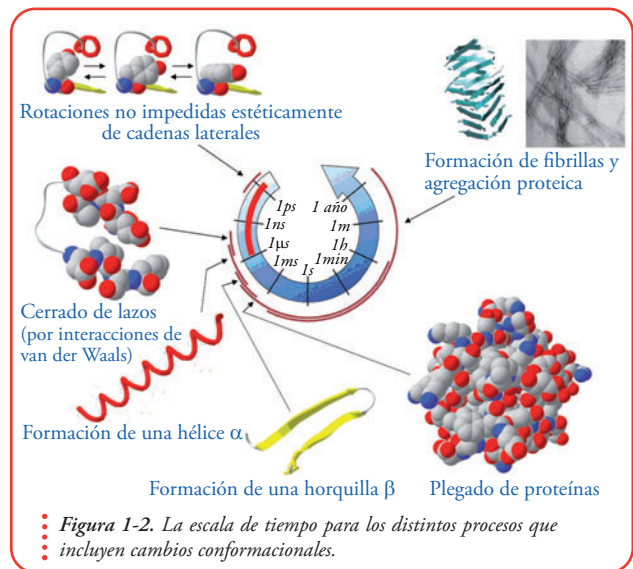
¿Cuánto mide una proteína? ¿En qué rango de tiempo ocurren los procesos relacionados con los cambios conformacionales?

Estas son dos preguntas importantes. Aunque más adelante nos preguntaremos cómo podemos hacer para medir el tamaño de una proteína o la velocidad del proceso de plegado, por ahora, será más que importante entender que, por lo general, se trata de objetos relativamente pequeños.

¿Cuán pequeños?

La hemoglobina humana, la proteína que transporta oxígeno y dióxido de carbono en la sangre, por ejemplo, es relativamente esférica y tiene un radio de aproximadamente 35,0 Ångström o 0,0000000035 metros.

Con respecto a la escala de tiempo en la que transcurren los procesos que involucran cambios conformacionales, es importante notar que se trata de una escala muy expandida, que va desde los picosegundos (ps), como es el caso de movimientos de rotación a través de un enlace ($1\text{ps} = 0,000000000001 = 1 \times 10^{-12}\text{s}$), pasando por los microsegundos (μs , $1\mu\text{s} = 1 \times 10^{-6}\text{s}$): el plegado de las proteínas que se pliegan más rápidamente y la escala se extiende hasta los minutos, horas e incluso días con las proteínas de plegado más lento y con los procesos de agregación que incluyen, por ejemplo, la formación de fibrillas amiloides y que tienen tanta importancia en salud humana. En este último caso un gran número de moléculas interactúan formando estructuras complejas; la conformación de las proteínas que forman este tipo de agregados suele ser una conformación alternativa a la nativa (Figura 1-2).



11

Representación de la estructura proteica (Primera Parte)

Por ahora no nos hemos detenido a analizar cómo representamos estructuras tan complejas como las de las proteínas, que pueden estar formadas por miles de átomos. Existen varias formas de hacerlo. Dependiendo de qué tipo de análisis queramos hacer

nos convendrá una forma u otra.

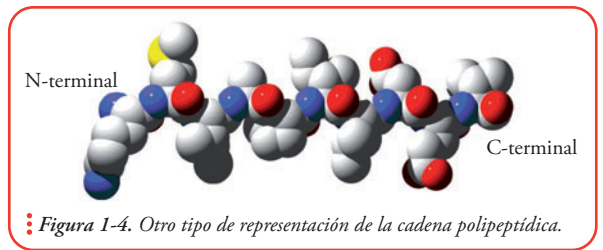
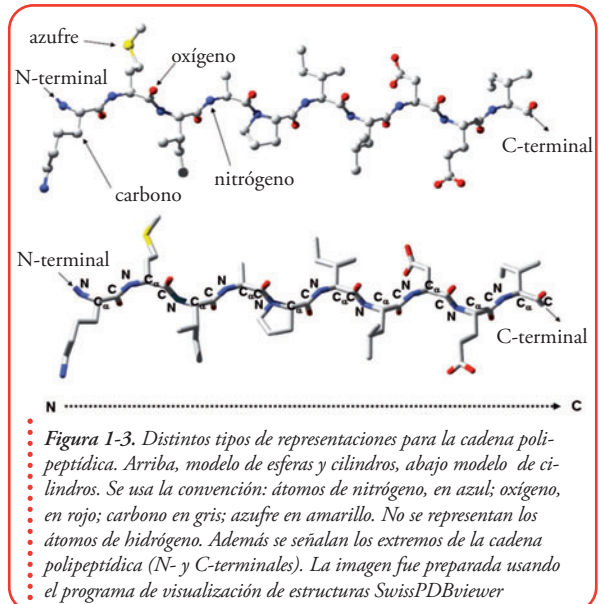
Necesitaremos un sistema de coordenadas para darle posición a cada átomo en el espacio, independientemente, del tipo de representación. El sistema de coordenadas más comúnmente usado es el cartesiano, donde para cada átomo tendremos coordenadas x , y , z .

Por lo general cuando representemos a los átomos, por ejemplo, a cada átomo como una esfera y cada enlace covalente como un cilindro, tendremos que utilizar un código de color para poder diferenciar cada tipo de átomo. Casi siempre se usa una convención en la que los átomos de nitrógeno se colorean en azul; los de oxígeno, en rojo; los átomos de carbono en gris, los átomos de azufre en amarillo y los átomos de hidrógeno, que con el objetivo de simplificar los modelos no siempre se representan, en blanco o gris claro o celeste (Figura 1-3), no se muestran los átomos de hidrógeno. Dependiendo del color con que se visualice el fondo esta convención puede modificarse. En general, deberíamos aclarar el código de colores que se usa.

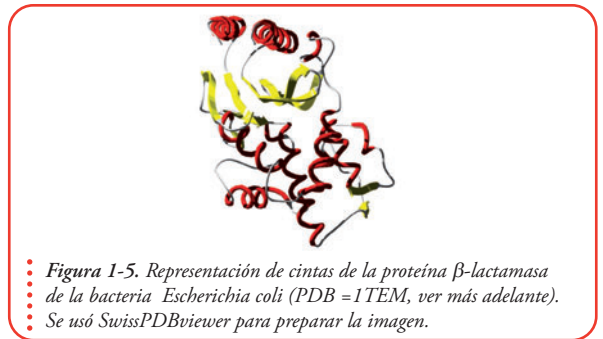
Este tipo de convenciones nos permite recorrer la cadena principal y reconocer muy fácilmente el sentido N-, C- terminal del esqueleto. Si bien a fines prácticos por ahora nos alcanza con este tipo de representación, es importante notar que estas esferas **no representan** los volúmenes reales de los átomos.

Más adelante recurriremos a modelos más realistas en los que representaremos los volúmenes relativos de cada uno de los átomos (Figura 1-4); estos modelos, más complejos, serán muy útiles, por ejemplo, cuando nos interese identificar el grado en el que los átomos están empaquetados o identificar cavidades, huecos o espacios vacíos en el interior de la matriz proteica.

También podemos representar el recorrido de la cadena polipeptídica en el espacio con modelos de cintas sin incluir un detalle a nivel atómico. Este tipo de modelo es muy útil para identificar en forma rápida la topo-



En la Figura 1-4 se representan los volúmenes relativos de cada átomo. Se usa la convención: átomos de nitrógeno, en azul; oxígeno, en rojo; carbono en gris; azufre en amarillo. No se representan los átomos de hidrógeno. Además se señalan los extremos de la cadena polipeptídica (N- y C-terminales). La imagen fue preparada usando el programa de visualización de estructuras SwissPDBviewer.



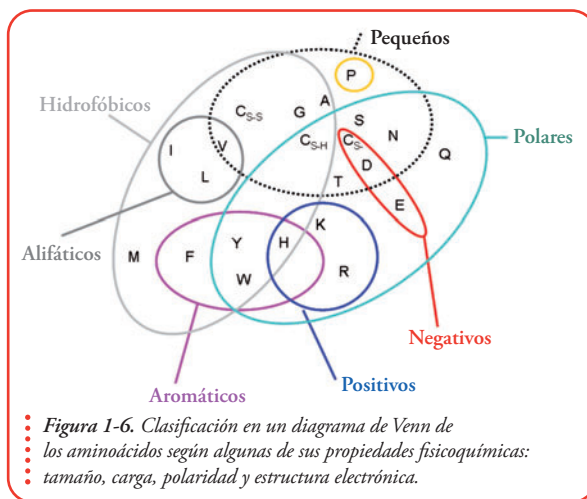
logía general de la macromolécula. Con este tipo de modelos podemos recorrer toda la proteína de extremo a extremo con relativa facilidad. En la **Figura 1-5** se muestra una representación de cintas de la β -lactamasa de la bacteria *Escherichia coli*, una enzima que degrada la penicilina G, un antibiótico muy comúnmente utilizado.

Los aminoácidos y sus propiedades

Los veinte aminoácidos naturales

Los aminoácidos que son utilizados por las células en la biosíntesis ribosomal de proteínas tienen en común un átomo de carbono central, el carbono alfa ($C\alpha$). Los aminoácidos que forman parte de las proteínas son alfa aminoácidos porque el $C\alpha$ está unido a un grupo amino (NH_2) y a un grupo ácido o carboxílico ($COOH$). El $C\alpha$ también está unido a un hidrógeno y a través de su cuarta valencia a un cuarto grupo que es variable. El código genético especificará 20 tipos diferentes de grupos, conocidos también como cadenas laterales (**Figura 1-6**).

Estos 20 grupos químicos definen entonces a los 20 tipos distintos de aminoácidos. Así, los aminoácidos se pueden agrupar en función de características generales de la estructura química de las cadenas laterales. Por ejemplo, podemos establecer categorías como cadenas laterales grandes y voluminosas y cadenas laterales pequeñas; cadenas laterales hidrofóbicas (apolares, ver **Tabla 1**) y cadenas laterales polares, entre estas últimas también podemos distinguir cadenas laterales polares cargadas y no cargadas. Asimismo podemos diferenciar aminoácidos con cadenas laterales aromáticas (los aminoácidos que tienen orbitales moleculares con electrones deslocalizados) de aminoácidos con cadenas laterales alifáticas. Algunos aminoácidos son claramente duales, como la cisteína (Cys) que, dependiendo del entorno químico, se puede comportar como polar e incluso cargado en su forma tiolato ($-CH_2-S^-$) mientras que en otros entornos pueden ser bien hidrofóbico, el grupo tiol ($-SH$) tiene menor capacidad de formar puentes de hidrógeno que el grupo alcohol ($-OH$), por ejemplo presente en la serina; particularmente, cuando las cisteínas forma parte de un puente disulfuro ($-S-S-$) el grupo químico se vuelve apolar. Otro ejemplo de este tipo de aminoácidos, es la histidina (His) que dependiendo de variaciones pequeñas de pH en el rangos de pH fisiológicos puede cargarse positivamente o encontrarse en su forma neutra triptofanos (Trp) y tirosinas (Tyr) en cierto sentido también pueden ser considerados aminoácidos duales, porque tanto el grupo $-OH$ de la cadena lateral de la tirosina como el nitrógeno del anillo aromático del triptofano pueden formar puentes de hidrógeno, pero también estos residuos pueden formar parte de regiones, extremadamente, apolares en el corazón de las proteínas. A su vez, los residuos cargados pueden tener carga



positiva como las lisinas (Lys) y las argininas (Arg) o negativa como los ácidos glutámico (Glu) y aspártico (Asp).

En una primera aproximación podemos hacer un intento por agruparlos, con las precauciones del caso:

1. Los apolares o hidrofóbicos que, como vamos a ver, predominan en el interior de las proteínas: valina (Val), leucina (Leu), isoleucina (Ile), fenilalanina (Phe), metionina (Met), tirosina (Tyr), triptófano (Trp), alanina (Ala).
2. Los aminoácidos cargados, que suelen estar presentes en la superficie de las proteínas expuestos al solvente acuoso: Lys, Arg, Asp y Glu. Sin embargo la parte hidrofóbica de cadena lateral de la Lys puede mantener, simultáneamente, interacciones con residuos apolares.
3. Los aminoácidos polares no cargados son serina (Ser), treonina (tha), asparragina (Asn), glutamina (Gln). Además la glicina (Gly), cuya cadena lateral es solamente un átomo de hidrógeno; esto como vamos a ver le otorga a este aminoácido propiedades muy particulares. La prolina (Pro) que es un iminoácido (es cíclico y con propiedades también muy particulares).

Si bien algunas cadenas laterales son muy parecidas, queda bien claro que es muy complicado agrupar los aminoácidos de una única forma. Esto ocurre por la inmensa gama de características fisicoquímicas que presentan en conjunto. Cada cadena lateral merece ser estudiada en forma particular.

Como ejemplo de escala de hidrofobicidad de los aminoácidos se muestra la escala de Eisenberg D., Schwarz E., Komarony M.y Wall R. publicada en la revista J. Mol. Biol. 179:125-142(1984).

Aminoácido. Valor de hidrofobicidad:

Ala	0,620	Gln	-0,850	Leu	1,060	Ser	-0,180
Arg	-2,530	Glu	-0,740	Lys	-1,500	Thr	-0,050
Asn	-0,780	Gly	0,480	Met	0,640	Trp	0,810
Asp	-0,900	His	-0,400	Phe	1,190	Tyr	0,260
Cys	0,290	Ile	1,380	Pro	0,120	Val	1,080

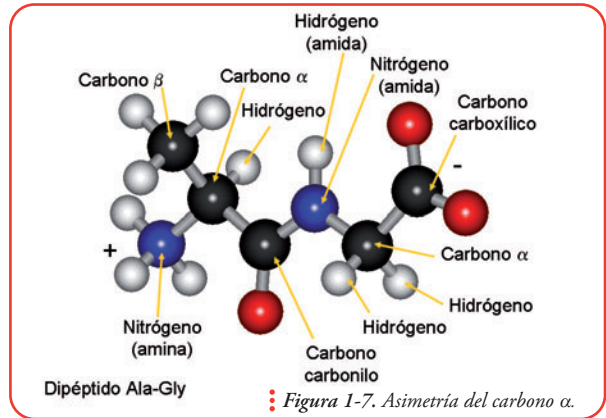
Tabla 1-1. Clasificación de los aminoácidos por su hidrofobicidad.

Los cuatro sustituyentes químicos unidos al C α son diferentes, por esto, los aminoácidos son moléculas quirales, asimétricas (**Figura 1-7**). Esto ocurre para todos con excepción de la Glicina que tiene 2 hidrógenos (el C α de la Gly no es un carbono quiral).

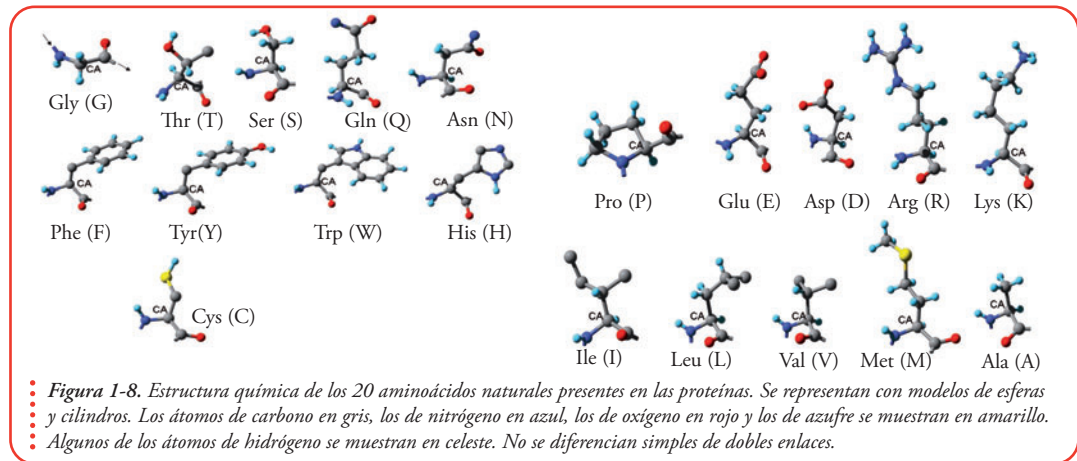
Los 19 restantes pueden ser de la forma L o D. Es importante que tengan en cuenta que la maquinaria de traducción para la biosíntesis de proteína reconoce, específicamente, como sustrato en la biosíntesis a los aminoácidos de la forma quiral L. Sin embargo en el laboratorio, mediante el uso de técnicas de síntesis química de péptidos podemos construir péptidos con aminoácidos de la forma quiral D e incluso péptidos en los que podemos mezclar aminoácidos D y L.

En algunos casos, ciertas cadenas laterales pueden modificarse con posterioridad a la síntesis ribosomal (el proceso de síntesis ribosomal de proteínas se conoce con el nombre de traducción). Estas modificaciones se conocen como modificaciones post-traduccionales. Hidroxilaciones, fosforilaciones, sulfataciones, acetilaciones, metilaciones, son algunas de las modificaciones postraduccionales más comunes.

Estas modificaciones en algunos casos funcionan como señales químicas en procesos biológicos alterando la interacción entre proteínas (proteína-proteína) o con otras macromoléculas (proteína-DNA o proteína-RNA, por ejemplo).



En la Figura 1-7 se representa el dipéptido Ala-Gly. Los cuatro sustituyentes químicos unidos al $C\alpha$ son diferentes: la cadena lateral (en el caso de la alanina es el grupo químico $-CH_3$), un átomo de hidrógeno ($-H$), el carbono carbonilo ($-C=O$) y el nitrógeno amida $N-H$ (o amina NH_3^+ , si se trata del residuo ubicado en el extremo N-terminal de la cadena polipeptídica). Únicamente en el caso de la glicina el carbono no es asimétrico.



El enlace peptídico

En las proteínas los aminoácidos están unidos mediante uniones peptídicas. Estas uniones tienen propiedades que determinan muchas de las características generales de estas macromoléculas.

El grupo carboxilo ($-COOH$) del primer aminoácido se condensa con el grupo amino ($-NH_2$) del siguiente aminoácido (eliminando una molécula de agua) formando un enlace amida (Figura 1-9).

Este proceso es repetido muchas veces en la etapa de elongación durante síntesis ribosomal. Así, la cadena polipeptídica se va alargando. La llamamos cadena polipeptídica porque, como dijimos antes, la podemos describir como un polímero de enlaces peptídicos. Esta cadena de enlaces peptídicos se conoce como esqueleto o cadena principal.

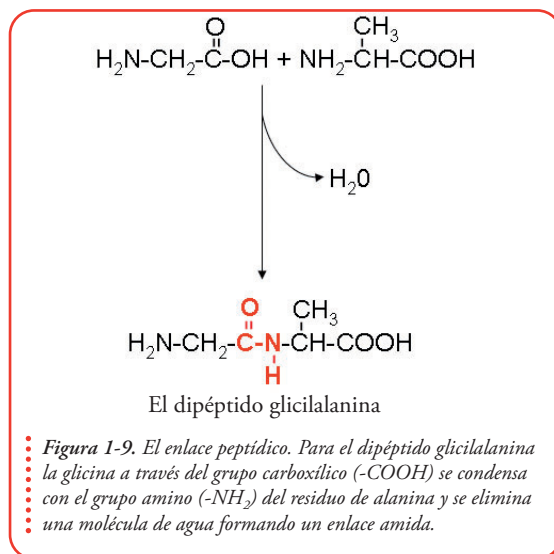
Es muy importante entender qué átomos forman parte del esqueleto polipeptídico (o cadena principal): el grupo NH (nitrógeno amida), el C α , (CA en la nomenclatura PDB) y el carbono carbonilo, C'=O, (el átomo C' unido al C α , al que también identificaremos de acuerdo con la nomenclatura PDB con una C); el átomo de carbono carbonílico está simultáneamente unido al NH (nitrógeno amida) del residuo siguiente (la unidad básica es entonces NH-C α H-C'O, ver **Figura 1-11**).

También es importante ver que el NH del primer aminoácido no participa de ningún enlace amida y, entonces, está en su forma -NH $_3^+$, este grupo se conoce amino terminal (químicamente es un grupo amina) o simplemente extremo N-terminal, a pH fisiológico le aportará una carga positiva a la proteína. Del mismo modo, el C'=O del último aminoácido forma parte del grupo C'OO $^-$ y es conocido como carboxilato terminal o más comúnmente como el extremo C-terminal de la proteína, a la que por lo general, (salvo a pHs extremadamente ácidos) le aportará una carga negativa a la proteína.

Una de las propiedades más importantes del enlace peptídico es su planaridad. El enlace peptídico tiene un 40% de carácter de doble enlace y como resultado la rotación a través de este enlace está impedida. Los seis átomos involucrados tienen una muy fuerte tendencia a ser coplanares. El carácter de doble enlace puede inferirse de la longitud del enlace peptídico (C'-N) que es de 1,33 Å, mientras que la longitud de un enlace típico C-N es de 1,45 Å (el caso de la unión C α -N). También puede verse que tampoco se trata de un doble enlace: la distancia entre el átomo de C y el de N en un doble enlace promedio (C=N) es mucho más corta, 1,25 Å.

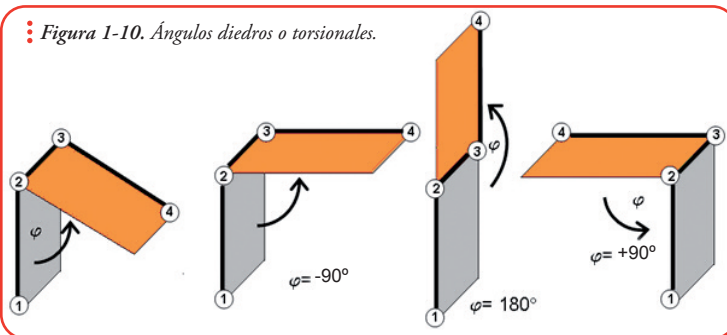
Así, en un polipéptido el enlace peptídico determina la formación de unidades planas repetitivas con mínima rotación a través del enlace C'-N. Estas unidades, por lo general, se encuentran en la configuración *trans*.

Al ángulo diedro definido por los cuatro átomos C $\alpha_{(i)}$ -C'-N-C $\alpha_{(i+1)}$ se lo conoce con el nombre de ω (**Figura 1-11**). Si bien la libre



*Para comprender el significado del término *trans* podemos pensar en la rotación a través de un enlace y en los ángulos torsionales. Un ángulo torsional (o diedro, **Figura 1-10**) es el ángulo entre dos planos a través de una recta que es la resultante de la intersección de ambos planos (podemos pensar en una bisagra, o en una puerta que se abre y que se cierra y el ángulo que forma con la pared), se define entonces con cuatro puntos en el espacio.*

Figura 1-10. Ángulos diedros o torsionales.



En la Figura 1-10 se representan cuatro átomos (1, 2, 3 y 4) hipotéticos unidos a través de enlaces covalentes. Se observa el ángulo φ entre los dos semiplanos, uno en color gris y otro en naranja. La recta resultante de la intersección de ambos semiplanos se superpone con el enlace entre los átomos 2 y 3. Cuando el valor de $\varphi=180^\circ$, los átomos 1 y 4 quedan en posición *trans*.

rotación a través del enlace C'-N está suprimida por el carácter de doble enlace: el ángulo torsional ω puede tomar dos valores 0° (*cis*) o 180° (*trans*). Estos valores pueden desviarse unos pocos grados sin afectar drásticamente la estabilidad imprimiendo cierta desviación con respecto a la planaridad ideal.

En la Figura 1-11 se muestran dos residuos consecutivos de una proteína humana, la frataxina. Se ve cómo los seis átomos C α , C', O del residuo Leu103 y N, H y C α , del residuo siguiente (Asp104) están ubicados en el mismo plano.

La configuración *trans* es, desde un punto de vista energético, unas mil veces más favorable que la configuración *cis* porque existe menos repulsión entre los átomos. En la configuración *trans* los átomos C $\alpha_{(i)}$ y C $\alpha_{(i+1)}$ (donde (i+1) es el residuo siguiente al residuo i) están a una distancia de 3,8 Å, mientras que en la configuración *cis* están a solamente 3,63 Å. En el caso de la configuración *cis*, los átomos de los residuos quedan demasiado cerca.

Además es importante ver que la presencia de un enlace peptídico en la configuración *cis* impone un fuerte cambio en la dirección que toma la cadena polipeptídica. Sin embargo, la configuración *cis* del enlace peptídico no se ve tan desfavorecida cuando el residuo siguiente al enlace en cuestión es una prolina (enlaces x-pro). En este caso el enlace peptídico entre un residuo y la prolina no tiene carácter de doble

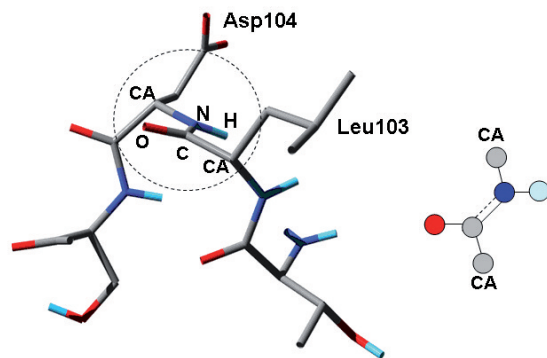


Figura 1-11. El ángulo diedro ω . En el enlace peptídico, ω está definido por los átomos C $\alpha_{(i)}$ -C'-N-C $\alpha_{(i+1)}$. También se lo representa con un esquema sencillo a la derecha. Se usa indistintamente C α y CA para referirse a los carbonos α . Se usó SwissPDBviewer para preparar la imagen.

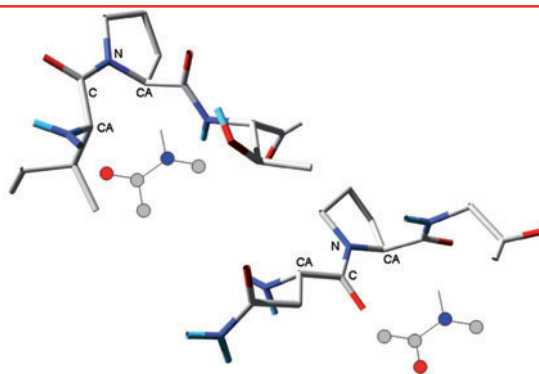


Figura 1-12. El ángulo diedro ω en el caso de un enlace peptídico entre los residuos X-Pro donde X es un aminoácido y Pro es prolina puede adoptar dos valores $\omega=0^\circ$, (configuración *cis*, ejemplo de la izquierda) o $\omega=180^\circ$ (configuración *trans*, ejemplo de la derecha). Se usó SwissPDBviewer para preparar la imagen y la estructura PDB 2TRX.

enlace. A su vez, el carácter cíclico de la Pro disminuye la energía ya que hay menos repulsión. La energía de las dos configuraciones ahora se vuelve comparable, como consecuencia podemos encontrar ángulos ω de 0° para enlaces x-pro en la estructura de ciertas proteínas (**Figura 1-12**).

El cambio entre configuración *cis* y *trans* es lento, ocurre en una escala de tiempo que va desde los 10 a los 100 segundos. Ambos estados están separados por una barrera energética muy elevada (cercana a las 20 kcal/mol), esta barrera puede disminuir en presencia de enzimas (proteínas conocidas como peptidil-prolil isomerasas) que disminuyen la energía del estado de transición de la reacción. Más adelante vamos a volver a discutir el problema de las prolinas, ya que la existencia de un enlace en configuración *cis* en la estructura de una proteína puede disminuir la velocidad de plegado considerablemente.

Como conclusión, la estructura electrónica del enlace peptídico restringe el número de conformaciones posibles que puede adoptar la cadena principal de una proteína.

Representación de la estructura proteica (segunda parte: archivos PDB)

Como ya hemos mencionado existe en la actualidad una base de datos muy importante de modelos experimentales de estructuras de proteínas: el PDB (*de Protein Data Bank en inglés*) o Banco de Datos de Proteínas. En este banco de datos se almacenan, entre otras cosas, las coordenadas espaciales de los átomos de cada modelo. Estos datos se depositan en forma de archivos que reciben el nombre de **archivos PDB**. Por ejemplo, en 1EKG.pdb está archivada la estructura de la frataxina humana. Cada uno de estos archivos tiene un nombre particular que permite identificar y no confundir un modelo con otro. El nombre es único. En estos archivos, además de las coordenadas atómicas, se almacena muchísima información útil para los investigadores acerca de los experimentos que dieron origen al modelo e información sobre el modelo en sí. Con una simple operación podemos acceder a estos archivos y no sólo ver los modelos, sino trabajar con ellos. Para esto se usan programas de computación especiales. Muchos de estos programas son gratuitos, muy buenos y fáciles de usar. Algunos de estos programas son Rasmol, SwissPDBviewer, MOL-MOL y VMD, entre otros. El uso de estos archivos permite, en muchos casos, elucubrar hipótesis para realizar nuevos experimentos, como por ejemplo iniciar el diseño de nuevos medicamentos (ver más adelante).

Los archivos PDB tienen una estructura particular que es importante comprender. Pero por ahora nos enfocaremos simplemente en la región del archivo que incluye las coordenadas espaciales de cada átomo pesado (en este caso denominaremos átomos pesados a los átomos que no son átomos de hidrógeno, estos son: nitrógeno, carbono, oxígeno y azufre, N, C, O y S, respectivamente). Como ejemplo se muestran a continuación las líneas correspondientes a los primeros 16 átomos del modelo construido por cristalografía y difracción de rayos X la frataxina (1EKG):

1	2	3	4	5	6	7	8	9	10	11	12
ATOM	1	N	LEU	A	90	1,020	13,240	5,227	1.00	37,83	N
ATOM	2	CA	LEU	A	90	-0,329	12,612	5,341	1.00	37,74	C
ATOM	3	C	LEU	A	90	-1,386	13,607	4,894	1.00	35,93	C
ATOM	4	O	LEU	A	90	-1,081	14,771	4,628	1.00	36,75	O
ATOM	5	CB	LEU	A	90	-0,633	12,224	6,794	1.00	39,88	C
ATOM	6	CG	LEU	A	90	0,491	11,857	7,765	1.00	40,72	C
ATOM	7	CD1	LEU	A	90	1,365	10,769	7,161	1.00	42,11	C
ATOM	8	CD2	LEU	A	90	1,305	13,102	8,082	1.00	41,63	C
ATOM	9	N	ASP	A	91	-2,630	13,147	4,809	1.00	33,80	N
ATOM	10	CA	ASP	A	91	-3,725	14,027	4,436	1.00	31,98	C
ATOM	11	C	ASP	A	91	-4,099	14,771	5,713	1.00	30,82	C
ATOM	12	O	ASP	A	91	-3,594	14,442	6,790	1.00	29,10	O
ATOM	13	CB	ASP	A	91	-4,913	13,227	3,895	1.00	33,03	C
ATOM	14	CG	ASP	A	91	-5,405	12,176	4,864	1.00	35,00	C
ATOM	15	OD1	ASP	A	91	-6,415	11,517	4,551	1.00	38,57	O
ATOM	16	OD2	ASP	A	91	-4,787	12,001	5,931	1.00	37,11	O

La **primer columna** (1) indica si se trata de un átomo (ATOM) que forma parte de la cadena polipeptídica o si se trata, por el contrario, de un átomo que forma parte de un ligando (como por ejemplo un inhibidor en una enzima), o tan solo moléculas de agua o átomos metálicos como hierro, cobre, zinc entre muchos otros.

En este caso veremos el término HETATM, como ejemplo se muestran las últimas 5 filas del archivo PDB que corresponden a cinco moléculas de agua.

HETATM	1018	O	HOH	77	-6,747	10,948	22,831	1.00	45,81	0
HETATM	1019	O	HOH	78	17,988	25,834	15,333	1.00	49,27	0
HETATM	1020	O	HOH	79	-9,053	19,621	24,657	1.00	44,44	0
HETATM	1021	O	HOH	80	13,025	9,676	19,422	1.00	40,88	0
HETATM	1022	O	HOH	81	5,559	12,413	14,018	1.00	38,16	0

Aquí, la **segunda** (2) columna indica el número de átomo y permitirá identificarlo; en el primer caso va del átomo 1 al 16 porque se muestra sólo una parte del archivo PDB, la correspondiente a los primeros 16 átomos que a su vez corresponden a los primeros dos residuos de aminoácidos (ver más adelante). En el segundo caso, al incorporar el resto de los átomos de frataxina y las moléculas de agua del modelo llegamos a la fila 1022 (se trata justamente de 1022 átomos). El modelo de la proteína (la frataxina propiamente dicha, sin tener en cuenta las moléculas de agua) incluye 941 átomos. Hay que tener en cuenta que los átomos de hidrógeno no



se incluyen en este modelo. Más adelante veremos en detalle por qué no se los incluye.

En la **tercera columna (3)** se describe de qué átomo se trata, con una nomenclatura que es la nomenclatura PDB: N es el nitrógeno del grupo amida, CA es el carbono alfa, C es el carbono carbonilo, O es el oxígeno carbonílico. CB es el carbono beta, CG carbono gama y CD1 y CD2 son los carbonos delta 1 y 2. En la **Figura 1-13** se muestra esta nomenclatura detalladamente para cada residuo. En el caso de las filas correspondientes a moléculas de agua, O es el oxígeno de cada molécula (HETATM 1018 O).

La **siguiente columna (4)** indica el tipo de residuo (Leu, por ejemplo) en código de tres letras. Se puede ver que los primeros 16 átomos corresponden a los primeros dos residuos de aminoácidos en la proteína: Leu, Asp.

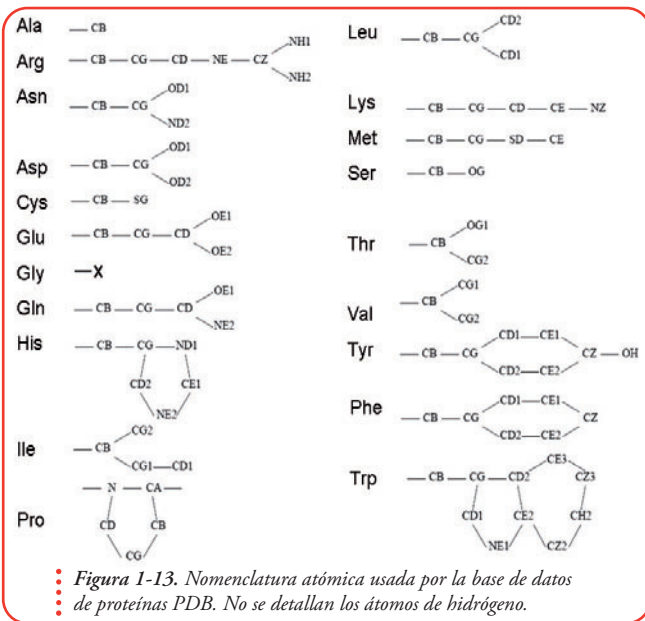
Luego, (**columna 5**) como en estos archivos, por diferentes motivos, puede haber más de una cadena polipeptídica, se anota a qué cadena corresponde el residuo mencionado en la columna anterior (en este caso a la cadena A).

En la **siguiente columna (6)** se anota el número de residuo. Este número (90 en este caso) suele estar relacionado con la posición de este residuo en la secuencia del gen de la proteína, en el caso de la frataxina, la proteína se sintetiza con un péptido señal N-terminal que en un proceso de maduración de la proteína es eliminado; el modelo 1EKG es un modelo que abarca desde el residuo 90 hasta el residuo 208 de la proteína.

Las **siguientes tres columnas (7, 8 y 9)** son muy importantes. Se trata de las coordenadas x, y, z. Los valores 1,020, 13,240, 5,227 ubican al nitrógeno (N) de la Leu 90 en una posición determinada en el espacio. De la misma manera, los valores -0,329, 12,612, 5,341 posicionan al carbono alfa (C α o CA en la convención PDB).

La próxima **columna (10)** es el factor de ocupación. Si un átomo pudiera ocupar, alternativamente, una de dos posiciones y cada una con una probabilidad del 50% (el caso en que ambas posiciones isoenergéticas) el factor de ocupación para ese átomo (o conjunto de átomos) tomaría el valor de 0,5 y no de 1 (probabilidad del 100%).

La **antéúltima, la columna (11)** es el factor B o factor de temperatura. El factor B está relacionado con la incertidumbre que tenemos en relación a la ubicación espacial del átomo. Este factor está relacionado por lo tanto con la movilidad o flexibilidad de la posición, pero también con la calidad del cristal que dio origen al experimento (ver



cristalografía y difracción de rayos X). Si se trata de una región sumamente rígida de la proteína los valores de los factores B para los átomos que componen esa región deberían ser bajos con respecto a otras regiones más móviles.

La última **columna** (12) indica el tipo de átomo, en términos generales (N, nitrógeno en la última columna y en la primera fila).

Libertad conformacional, ángulos torsionales

A pesar del carácter rígido que imprime el enlace peptídico al esqueleto proteico, la cadena principal tiene cierto grado de libertad de rotación en torno a dos enlaces: $C\alpha-C'$ y $N-C\alpha$ (Figura 1-14).

Esta libertad conformacional *es clave* en la adquisición de la estructura proteica. Al ángulo torsional que describen los átomos $C'-N-C\alpha-C'$ (a través del enlace $N-C\alpha$) se lo llama **phi** (ϕ) y al ángulo torsional que conforman los átomos $N-C\alpha-C'-N$ (a través del enlace $C\alpha-C'$) se lo llama **psi** (ψ) (Figuras 1-14 y 1-15). De esta manera cada $C\alpha$, y, *en definitiva, cada residuo de la proteína*, está asociado con dos ángulos conformacionales, un **phi** y un **psi**. En la Figura 1-15 se muestra, en un esquema, cómo podemos identificar a los ángulos **phi** y **psi**. También, cómo quedan determinados los planos que formarán el ángulo diedro, en definitiva, cómo quedan determinadas las torsiones y cómo estas torsiones definen el recorrido de la cadena polipeptídica en el espacio.

Como estos son los únicos ángulos torsionales del esqueleto del polipéptido con libertad rotacional, una vez que cada par de ángulos queda definido para cada residuo, entonces queda, completamente, definido el recorrido de la cadena polipeptídica en el espacio.

La mayor parte de las combinaciones posi-

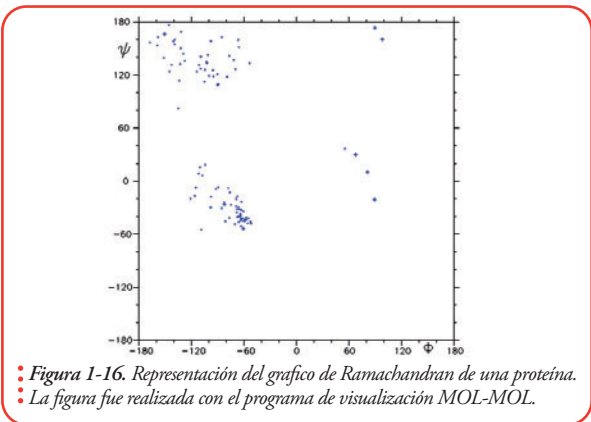
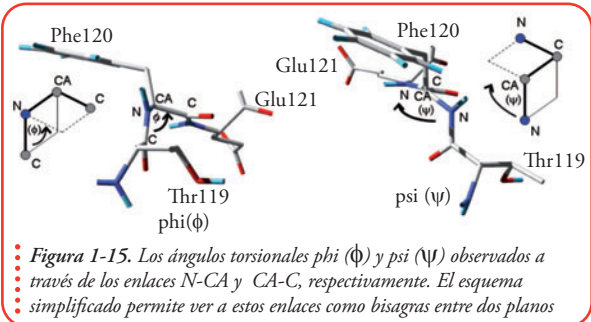
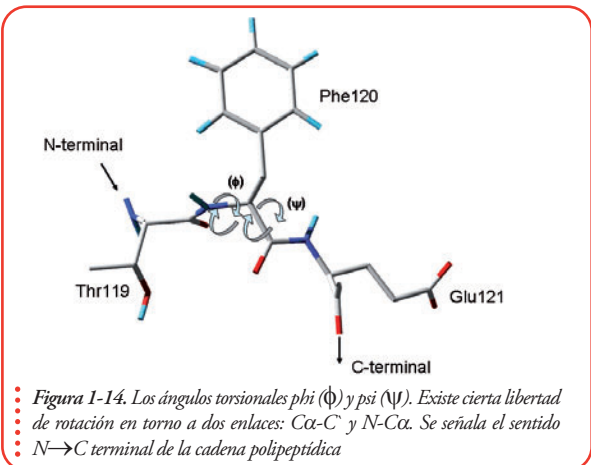


Gráfico de la Figura 1-16 conocido como Ramachandran. Se representa para cada aminoácido de la frataxina humana el par de ángulos diedros ϕ y ψ correspondiente. Para las glicinas particularmente se usó el símbolo (+).

bles de **phi** y **psi** para un aminoácido pueden considerarse muy poco probables o, directamente, imposibles debido a las colisiones intramoleculares que se producirían entre la cadena principal y las cadenas laterales en consecuencia a la elevadísima energía que poseerían estas conformaciones. Sobre la base de consideraciones geométricas puede calcularse cuáles son las conformaciones permitidas y cuáles son las conformaciones prohibidas. Los pares de ángulos **phi** y **psi** permitidos pueden, entonces, graficarse uno contra el otro. En la **Figura 1-16** se muestra el gráfico de Ramachandran de la frataxina humana (1EKG). Aquí, cada par de ángulos diedros (**phi** y **psi**) ha sido representado con un punto o una cruz azul.

En un modelo de poli-L-alanina (un polipéptido construido con alaninas), los C β podrían representar el mínimo de interacciones posibles entre las cadenas laterales de los aminoácidos comunes en las proteínas. Sin embargo, este modelo no describe correctamente el comportamiento observado para el aminoácido glicina ni tampoco para Pro. También resulta un modelo pobre en el caso de cadenas laterales muy voluminosas.

Ramachandran fue el científico que realizó los primeros cálculos sobre combinaciones permitidas. Estos cálculos publicados en 1963, los realizó en un contexto en el que se debatía, fuertemente, sobre la estereoquímica de la estructura nada más ni nada menos que del colágeno, que había sido modelada previamente (1954) por el y por G. Kartha.

Organización jerárquica y modular.

Estructura secundaria y motivos estructurales

Con la determinación por cristalografía y difracción de rayos-X de las primeras estructuras proteicas los proteinólogos aprendieron cosas muy importantes:

- a. el interior de las proteínas está compuesto, casi exclusivamente, por aminoácidos no polares y
- b. los aminoácidos están muy bien empaquetados.

Como consecuencia de ambos hechos puede deducirse que las proteínas tienen corazones hidrofóbicos y que su superficie es, predominantemente, polar. Existen algoritmos que nos permiten medir la superficie y la accesibilidad al solvente de los aminoácidos que componen una proteína. Esto se hace en base al modelo atómico de alta resolución de la proteína, es decir, cuán expuesta se encuentra tal o cual cadena lateral. Para esto se hace rotar una esfera de un radio determinado (por ejemplo 1.4 Å, en este caso simulando una molécula de agua) por sobre el modelo proteico y se determina con qué átomos contacta la esfera y con qué átomos no. Un punto clave es que la cadena principal posee grupos polares: el grupo amida (–NH) y el carbonílico (–C=O). El primero, un donador para la formación de puentes de hidrógeno y, el segundo, un aceptor: (–N H•••O=C–). Por lo tanto, estos grupos polares deben estar compensados entre sí en el interior, fuertemente, apolar de la macromolécula en su conformación nativa, correctamente plegada.

Con los primeros modelos de difracción de rayos X de alta resolución de estructuras de proteínas, también, se puso en evidencia la existencia de estructuras regulares o periódicas en las que se satisface la neutralización de estos grupos. Así se corroboraron los modelos de Linus Pauling para péptidos

construidos en base a la aplicación del conocimiento sobre la estructura de pequeñas moléculas por: la estabilización por puentes de hidrógeno intramoleculares de la conformación proteica, y la existencia de estructuras helicoidales. A estas estructuras periódicas (combinaciones periódicas de ángulos diedros ϕ y ψ) se las conoce con el nombre de *estructura secundaria*. Curiosamente, las proteínas están formadas, principalmente, por dos tipos de estructuras periódicas: hélices α y láminas β . Además, existen conectores que vinculan a estos elementos periódicos; así estos pueden disponerse espacialmente de diversas maneras.

Así podemos rever el gráfico de Ramachandran (mencionado con anterioridad) y asignar a cada zona de este gráfico alta o baja probabilidad a cada tipo de estructura periódica. En la Figura (ver numeración) se muestra con una β la región de máxima probabilidad para estructuras del tipo lámina β y con α la región de máxima probabilidad para estructuras de tipo helicoidales.

En el gráfico de Ramachandran, de una proteína podemos encontrar, fácilmente, aquellos residuos en los que se producen torsiones (combinaciones de ángulos diedros) atípicas (Figura 1-17). Como la cadena lateral de la glicina es un átomo de hidrógeno los ángulos torsionales de este aminoácido no están restringidos por el C_{β} , típicamente podremos encontrar residuos de Glicina por fuera de las regiones de máxima probabilidad. Por otro lado, la cadena lateral del residuo de Pro consiste en tres átomos de carbono, pero estos forman un anillo de cinco miembros con el C_{α} y con el N imida de la cadena principal. Como resultado, se produce una disminución considerable de la libertad conformacional; ϕ tiene un mínimo de energía en -60° y ψ tiene dos mínimos cerca de -55° y cerca de 145° .

Estructuras helicoidales

Las hélices alfa (α)

Las hélices α son fácilmente reconocibles en la estructura proteica (Figura 1-18, paneles A y C). Los ángulos torsionales ϕ (ϕ) y ψ (ψ) toman valores cercanos a -57° y -47° , respectivamente. Estas hélices tienen 3,6 residuos por vuelta y por cada aminoácido se produce un incremento de la hélice de 1,5 Å (traslación por residuo). En este tipo de elemento de estructura secundaria los átomos de la cadena principal quedan empaquetados muy eficientemente (interacciones de van der Waals favorables). Además, cada oxígeno carbonílico, por ejemplo el del residuo número (i) en la hélice, forma un puente de hidrógeno con el grupo amida (-N-H) del residuo (i+4). Así, quedan satisfechos todos los puentes de hidrógeno in-

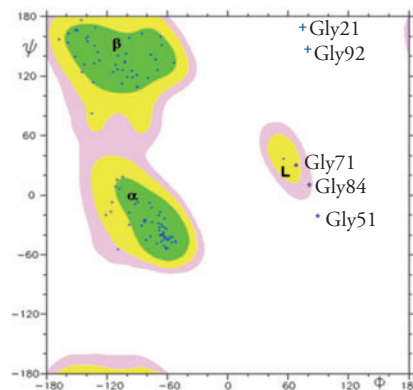


Figura 1-17. Combinaciones de ángulos diedros ϕ y ψ propias de estructuras periódicas tipo hélices α o hebras β .

En la Figura 1-17 nuevamente se representa para cada aminoácido de la frataxina humana el par de ángulos diedros ϕ y ψ correspondiente y para las glicinas particularmente se usó el símbolo (+). Se muestran las regiones de máxima probabilidad en verde y se señala el tipo de estructura secundaria.

La figura fue realizada con el programa de visualización Mol-Mol

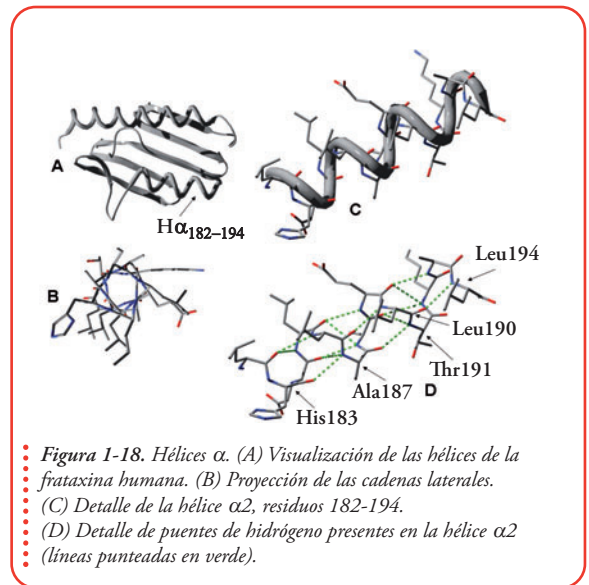


tramoleculares (Figura 1-18, Panel D).

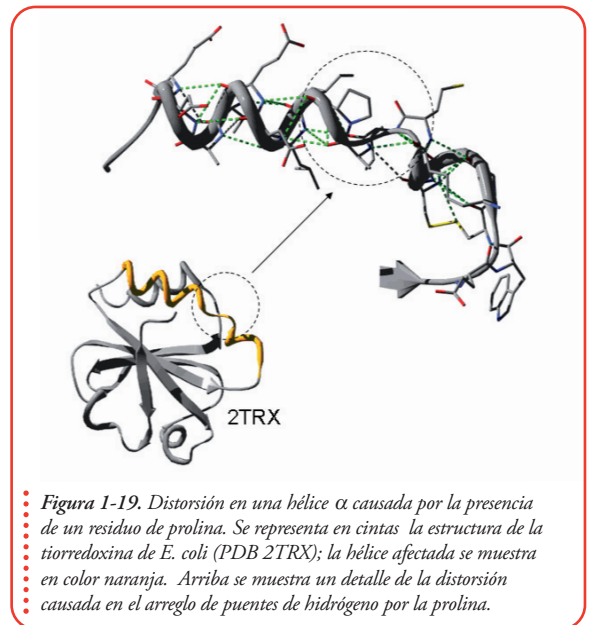
Es importante notar que las cadenas laterales se proyectan hacia fuera (Figura 1-18, Panel B) y también contribuyen con la estabilidad de la hélice a través de las interacciones que establecen entre sí, localmente, por ejemplo las interacciones de van der Waals (ver más adelante) que se establecen entre las cadenas laterales de los residuos (i) e (i+4).

Algunas conformaciones de las cadenas laterales (rotaméricas) quedan altamente desfavorecidas porque producen solapamiento con otros átomos de residuos vecinos de la hélice α . Los residuos que poseen un $C\beta$ ramificado (treonina, isoleucina y valina tienen mayores restricciones conformacionales. Por otro lado, los residuos con cadenas laterales polares (por ejemplo, serina y asparagina) pueden competir en la formación de puentes de hidrógeno con átomos de la cadena principal. Los residuos muy voluminosos en los extremos de hélices terminales (que presentan un trupo $-NH_3^+$ o un $-COO^-$) también suelen no ser convenientes estructuralmente porque desolvatan la carga, en estos casos se ven favorecidos residuos con cadena lateral pequeña. La introducción de glicinas en una hélice puede desestabilizarla significativamente, porque se le otorga a este elemento de estructura más libertad conformacional, y el péptido puede explorar otras conformaciones, por lo general menos probables (otros ángulos diedros phi (ϕ) y psi (ψ)). Es importante tener en cuenta que el residuo de prolina es, directamente, incompatible con la conformación helicoidal de tipo hélice α . En este caso, el último átomo de la cadena lateral está unido al nitrógeno de la cadena principal formando un anillo de cinco miembros ($C\alpha-CH_2-CH_2-CH_2-N$). Este hecho deter-

mina la ausencia de un grupo N-H que participe como donador en la formación de un puente de hidrógeno intrahelicoidal. Además la presencia de prolina en una hélice produce una torsión de la hélice y defectos de empaquetamiento de los átomos de la cadena principal. En la Figura 1-19, se muestra un modelo de cintas de la estructura de la tiorredoxina de *Escherichia Coli* PDB: 2TRX. Se puede ver cómo la prolina en este caso interrumpe la hélice



• Figura 1-18. Hélices α . (A) Visualización de las hélices de la frataxina humana. (B) Proyección de las cadenas laterales. (C) Detalle de la hélice α_2 , residuos 182-194. (D) Detalle de puentes de hidrógeno presentes en la hélice α_2 (líneas punteadas en verde).



• Figura 1-19. Distorsión en una hélice α causada por la presencia de un residuo de prolina. Se representa en cintas la estructura de la tiorredoxina de *E. coli* (PDB 2TRX); la hélice afectada se muestra en color naranja. Arriba se muestra un detalle de la distorsión causada en el arreglo de puentes de hidrógeno por la prolina.

(círculo en puntos negros) y cómo se interrumpe la conectividad interna de puentes de hidrógeno. A pesar de todo esto, se las puede encontrar como primer residuo de hélice, o formando parte de hélices de longitud considerable.

Como todos los puentes de hidrógeno intrahelicoidales se establecen con el mismo sentido y dirección quedando alineados con el eje longitudinal de la hélice, cada uno con un momento dipolar que proviene de que los grupos N-H y C=O poseen distinta polaridad. Estos dipolos alineados con el eje de la hélice forman un macro dipolo resultante de la sumatoria de cada uno de ellos (**Figura 1-20**).

Así, residuos con cadenas laterales cargadas negativamente (cadenas laterales de ácido glutámico o ácido aspártico) ubicadas en el extremo C-terminal de la hélice podrían desestabilizar la conformación helicoidal a través de la desestabilización del macro dipolo que ubica polaridad negativa en el extremo del elemento helicoidal. Del mismo modo son desestabilizantes las cargas positivas (residuos de lisina y arginina) ubicadas en el extremo N-terminal de la hélice. Por el contrario, residuos cargados, positivamente, en el extremo C-terminal o cargados negativamente en el extremo N-terminal de la hélice estabilizan el macro dipolo, y en consecuencia al elemento helicoidal.

En una proteína las hélices pueden localizarse en la superficie. En este caso una cara estará en contacto con el solvente y la otra estará en contacto con el interior apolar. En este tipo de hélices los residuos polares y apolares deben estar distribuidos en la secuencia periódicamente. A este tipo de hélices se las conoce con el nombre de hélices anfipáticas. Los péptidos que forman este tipo de hélices en el contexto de la proteína nativa aislados en solución (sin el resto de la proteína) requieren muchas veces cosolventes como trifluoroetanol o detergentes para adquirir el plegado helicoidal nativo. Esto se debe a que la superficie apolar que se genera en el proceso de plegado puede ser sumamente inestable en solución acuosa.

En otros casos los elementos helicoidales pueden estar, completamente, ocultos en el corazón hidrofóbico de la proteína o, completamente, expuestos al solvente. En el caso de muchas proteínas de membrana, las hélices están inmersas en un ambiente hidrofóbico. La mayor parte de las cadenas laterales serán por lo tanto apolares. Estas proteínas suelen tener varias secciones transmembrana. Tanto estas secciones como las hélices o hebras beta de *core* (del interior de la proteína) que son fuertemente apolares pueden decidirse a partir del perfil de hidrofobicidad de la proteína: un gráfico como el de la **Figura 1-21** (para la frataxina humana) en el que se grafica el índice de hidrofobicidad por residuo de aminoácido (ver **Tabla 1-1**) en función de la posición en la secuencia. Una vez que encontramos la secuencia en la base de datos (podemos buscarla en <http://www.ncbi.nlm.nih.gov/sites/entrez>) resta elegir la escala de hidrofobicidad, en este caso se eligió la escala de Eisenberg D., Schwarz

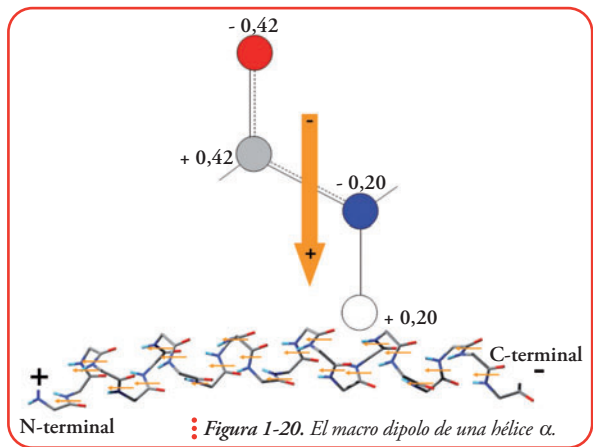


Figura 1-20. En el panel superior se muestra en un esquema cómo se produce el dipolo sobre la unidad peptídica. Se muestran las cargas parciales aproximadas sobre cada uno de los átomos. En el panel inferior de la figura se muestra una de las hélices α de la frataxina humana y cómo se alinearían los dipolos de las unidades peptídicas con el eje mayor de la hélice, en rojo oxígeno, en gris carbono, en azul nitrógeno y en blanco hidrógeno. Los valores de las cargas parciales son los sugeridos por Branden y Tooze.

E., Komarony M.y Wall R. publicada en la revista J. Mol. Biol. 179:125-142(1984) y se usó una aplicación disponible en la web (<http://ca.expasy.org/tools/protscale.html>). En esta escala como se mencionó, los valores más negativos corresponden a residuos cargados.

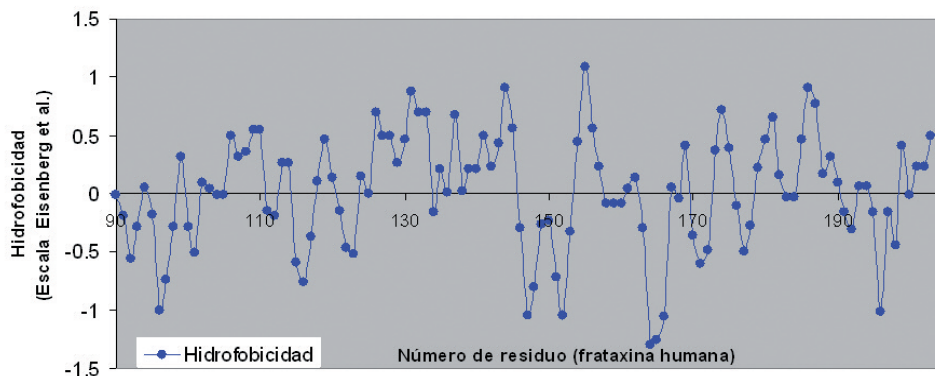


Figura 1-21. Perfil de hidrofobicidad de la frataxina humana. Se gráfica la hidrofobicidad en función del número de residuo.

En muchos casos puede ser muy útil la construcción de diagramas de tipo rueda o espiral. En estos gráficos (Figura 1-22) cada vuelta de hélice (3,6 residuos de aminoácidos) se grafica posicionando cada residuo cada 100° (una vuelta/3,6 residuos = $360^\circ/3,6=100$). Siempre hay que tener en cuenta que se trata de una hélice canónica (un modelo) y que las hélices reales pueden tener perturbaciones impuestas por la secuencia o por la estructura del resto de la proteína. Observando la proyección de los residuos puede verse, fácilmente, si una cara de la hélice, por ejemplo, es polar y la otra apolar, como en el caso de las hélices anfipáticas. En la Figura 1-22 se muestra el ejemplo de la melitina un péptido encontrado en altas concentraciones en el veneno de abejas y también se muestra una de las hélices de la frataxina. Los residuos apolares se muestran recuadrados y en azul, los residuos polares en fucsia.

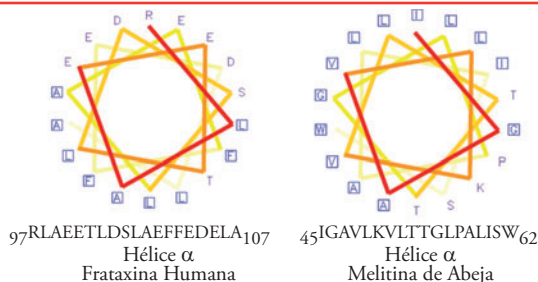


Figura 1-22. Construcción de diagramas de tipo rueda o espiral.

En la Figura 1-22 la proyección de los residuos permite identificar si una de las cara de la hélice es apolar; a la derecha la melitina y a la izquierda unas de las hélices de la frataxina. Los residuos apolares están recuadrados y en azul, los residuos polares en fucsia.

A partir de este tipo de gráficos podemos deducir cuál de las dos caras de la hélices α de la frataxina interacciona con el resto de la macromolécula (cara apolar) y cuál es la cara que interacciona con el solvente (cara polar). La melitina, por otro lado, forma un tetrámero (cuatro moléculas de melitina) en el que las superficies apolares están en contacto. También puede interaccionar con otras proteínas y con membranas biológicas.

Las hélices 3_{10} y las hélices π

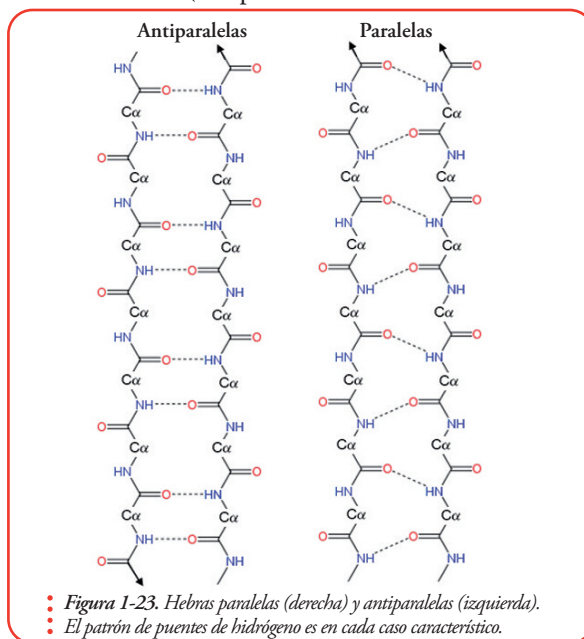
Existen otras conformaciones helicoidales en las que la cadena principal puede adoptar una conformación más laxa que una hélice α y con poco contacto entre sus átomos (con puentes de hidrógeno $(i, i+5)$), las hélices π , o menos laxa, las hélices 3_{10} (puentes de hidrógeno $(i, i+3)$). Esto quiere decir que el puente de hidrógeno intrahelicoidal que forma el residuo i no se establece con el residuo $i+4$, como en el caso de las hélices α , sino que se establece con el residuo $i+5$ o con el residuo $i+3$. En el segundo caso, quedan por vuelta de hélice tres residuos y diez átomos entre el átomo donador y el átomo aceptor que establecen el puente de hidrógeno (por este motivo se llaman hélices 3_{10}).

Las láminas beta (β)

En conjunto con las hélices α , el segundo elemento estructural regular, fácilmente, identificable en las estructuras proteicas son las láminas β . Estos elementos, en contraste con las hélices α , están compuestos por varias regiones de la cadena polipeptídica. Estas regiones poseen una estructura extendida con combinaciones de ángulos diedros **phi** (ϕ) y **psi** (ψ) que caen en la región permitida del cuadrante superior izquierdo en el gráfico de Ramachandran. Cada porción (cada hebra de la lámina), típicamente, está formada por 5-10 residuos con una periodicidad aproximada de 2 residuos (dos cadenas laterales consecutivas quedan a 180° , una hacia arriba y otra hacia abajo) con una traslación de $3,4 \text{ \AA}$ por residuo. Contrariamente, con la disposición de puentes de hidrógeno en las hélices (los puentes son intrahelicoidales), en las láminas los puentes de hidrógenos se forman entre hebras: entre el grupo amida ($-\text{NH}$) y el carbonílico ($-\text{C}=\text{O}$) de residuos de dos hebras distintas. Así, en el caso de las hojas o láminas β intramoleculares las interacciones entre hebras se dan entre residuos que están distantes en la estructura primaria. También puede tratarse de hebras que corresponden a distintas cadenas polipeptídicas (láminas beta intermoleculares).

Las hebras pueden interactuar, en principio, de dos maneras diferentes. Si ambas hebras alterna el sentido se tratará de hebras antiparalelas en cambio si corren con el mismo sentido (N- C-terminal) entonces se las conoce como hebras paralelas; (Figura 1-23).

Típicamente, para las hebras antiparalelas los ángulos diedros **phi** (ϕ) y **psi** (ψ) toman los valo-



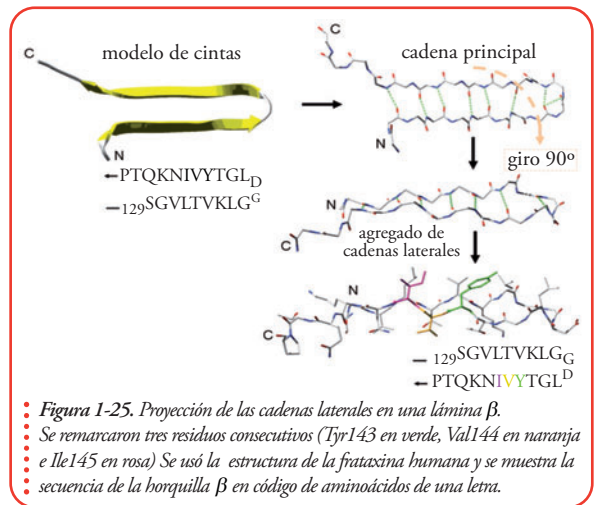
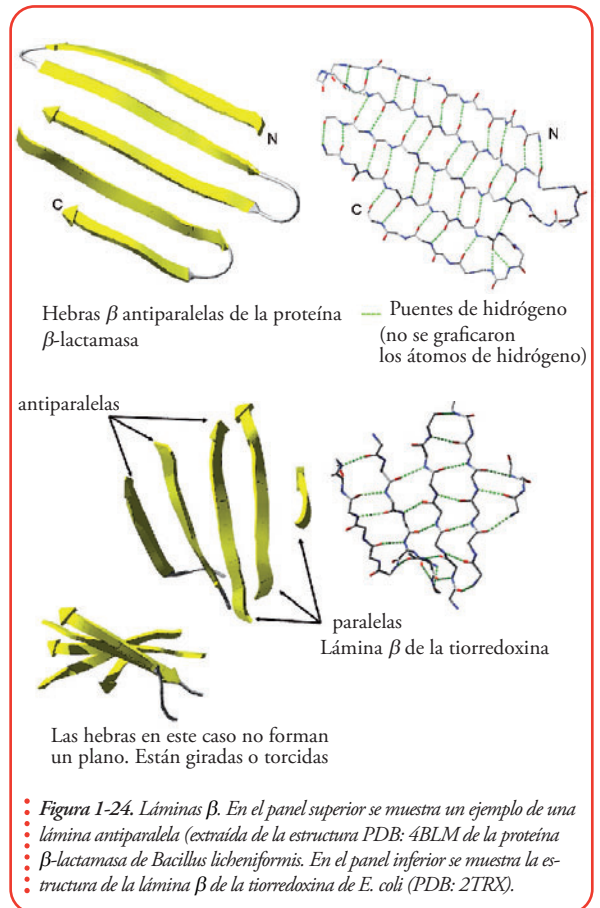
res de -139° , $+135^\circ$, respectivamente, y para las hebras paralelas toman valores de -119° y -113° .

En la **Figura 1-23** se puede observar el tipo de patrón de puentes de hidrógeno que se forma en cada caso. Es importante ver que en las **Figuras 1-23** y **1-24** no se muestran las cadenas laterales de los aminoácidos. Una lámina también puede estar formada por una mezcla de hebras paralelas y antiparalelas; a su vez, la lámina puede ser plana o no; si no es plana se la llama lámina u hoja torcida (*twisted* en inglés), este es el caso de la lámina de la tiorredoxina (**Figura 1-24**, panel inferior).

En la **Figura 1-25** se muestra un fragmento de la lámina β de la frataxina (residuos 129 a 150). Aquí es importante ver que, como se dijo anteriormente, las cadenas laterales consecutivas quedan a 180° , una hacia arriba de la lámina y otra hacia abajo. Si bien las láminas β se han catalogado como componentes de la estructura secundaria de las proteínas, debemos tener presente las grandes diferencias estructurales mencionadas más arriba entre estos elementos. Estas diferencias (por ejemplo, una lámina puede formarse por la interacción de dos hebras no consecutivas) tienen consecuencias importantes en el proceso de plegado.

Estructura super-secundaria y motivos estructurales

En las proteínas encontramos motivos formados por la asociación (y la interacción) de elementos de estructura secundaria. Estos motivos se conocen con el nombre de estructura supersecundaria y son patrones que se repiten con regularidad en las proteínas. Un ejemplo simple que comprende sólo elementos helicoidales y conectores son los paquetes de hélices (α - α , en inglés *helix bundles*), las hélices pueden estar unidas por



conectores cortos, incluso de tan sólo 3 residuos. Otro motivo de estructura supersecundaria es el motivo hélice lazo hélice (*helix-loop-helix* en inglés). En este caso dos hélices están conectadas por un lazo que puede variar en su longitud. Otro ejemplo son dos hebras beta antiparalelas unidas por un conector corto. A estos conectores se los llama β -turn (en inglés) o giro β en castellano (conector de dos residuos) y γ -turn o giro γ (conector de un único residuo). Se puede ver que estos conectores poseen secuencias que posibilitan este tipo de giro de la cadena polipeptídica. Hay dos tipos de vueltas β típicas en proteínas (β -turn de tipo I y II). En el caso de los de tipo II debe haber, inexorablemente, un residuo de Gly (recordar que la glicina no tiene carbono β y su cadena lateral es mínima, tan sólo un átomo de hidrógeno) así se evitan colisiones entre átomos de la cadena principal con átomos de este residuo. Además, en los giros β los residuos $i+1$ e $i+2$ (el residuo i es el inmediatamente anterior al giro) nunca forman puentes de hidrógeno y el plano que forma el enlace peptídico entre ambos residuos es perpendicular a la dirección que toma la cadena principal en las hebras β . En cambio los residuos i e $i+3$ participan en un puente de hidrógeno (en la Figura 1-26 se muestra un ejemplo de giro β).

Las hebras beta pueden, como se mencionó más arriba, formar hojas o láminas β y estas pueden tener una topología sencilla como en el caso de las horquillas β (en inglés, β -hairpins) y los arreglos tipo β meandros. Sin embargo, existen topologías u arreglos de hebras β más complejos con conectores más largos en forma de lazo (*loop* en inglés) Este tipo de arreglos permite acomodar, muy eficientemente, los átomos de la cadena principal, de tal forma que se establezcan, eficazmente, los puentes de hidrógeno entre hebras. La formación de puentes de hidrógeno intramoleculares estará muy favorecida en ambientes no polares como lo es el interior de una proteína. En el caso de que dos hebras fuesen paralelas, el conector debe ser, suficientemente, largo (tan largo como la hebra). Suele encontrarse un elemento α helicoidal cumpliendo ese rol. Así, tenemos este otro arreglo, el motivo β - α - β . En este caso ambas hebras forman puentes de hidrógeno entre hebras y están a su vez empaquetadas con la hélice.

Puentes disulfuro y tioles libres

Hemos dejado un poco de lado un punto que es realmente importante. La existencia de

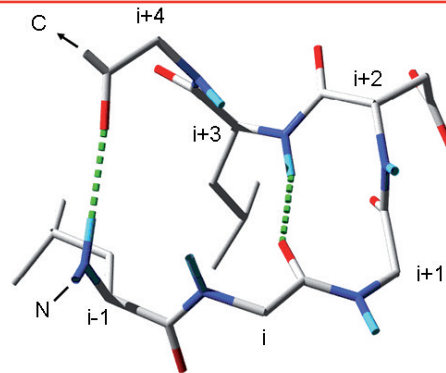


Figura 1-26. Conectores de hebras β . Se muestra un típico giro β . Se detallan los puentes de hidrógeno en líneas punteadas verdes y los extremos N- y C- terminales con una N y una C, respectivamente. Se nombran los residuos en función de su ubicación en la estructura del giro. $i-1$, i , $i+1$, $i+2$, $i+3$ e $i+4$.

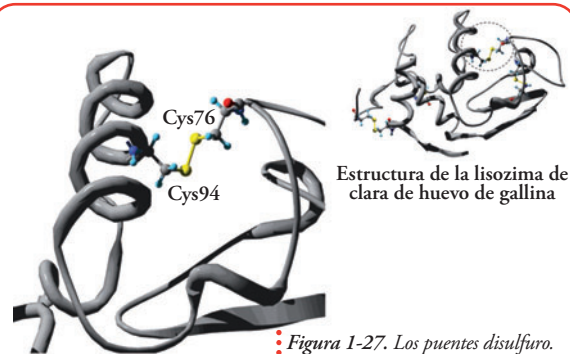


Figura 1-27. Los puentes disulfuro.

En la Figura 1-27 se muestra en detalle uno de los cuatro puentes disulfuro presentes en la lisozima de clara de huevo de gallina. A la derecha, la estructura completa de la proteína en la representación de cintas y con esferas y cilindros se representan los ocho residuos de cisteína. A la izquierda, se amplía la región ubicada dentro del círculo punteado de la imagen de la derecha y se muestra un detalle de las dos cisteínas involucradas en uno de los puentes disulfuro (Cys 76 y Cys 94). Se usó SwissPDBviewer para preparar la imagen.

puentes disulfuro en las proteínas. Dos cisteínas pueden oxidarse y formar un puente disulfuro. Esta oxidación puede ocurrir mediante la presencia de otro puente disulfuro o, alternativamente, puede ser catalizada por oxígeno y metales disueltos en la solución de proteína. Los puentes disulfuro son enlaces que restringen la movilidad de esa región de la proteína. Para que los puentes disulfuro se formen debe estar consolidada la estructura de la macromolécula de tal forma que ambas cisteínas queden ubicadas, correctamente, en el espacio (los dos átomos de azufre que forman un enlace disulfuro están separados por una distancia de unos 2Å). En la **Figura 1-27** se muestra como ejemplo uno de los cuatro puentes disulfuro de la proteína lisozima de clara de huevo de gallina.

Existen reactivos especiales para la reducción de puentes disulfuro (ruptura del enlace), uno de los más eficientes es el ditioneitol DTT y uno de los más comúnmente utilizados es el β mercaptoetanol.

En este contexto, vale la pena destacar que la reactividad de los residuos de cisteína aumenta a medida que el pH de la solución es más alcalino. Concretamente el $pK_{Cys} = 8,4$. Esto quiere decir que a $pH=8,4$ el 50% de las cisteínas debería estar en su forma aniónica (-S⁻), siendo este grupo, el tiolato, el grupo reactivo.

Otra reacción química de gran utilidad es la reacción de Ellman o del ácido 5,5'-Ditio-bis(2-nitrobenzico). Esta reacción se usa para cuantificar tioles libres en las proteínas, es decir cisteínas que no forman parte de puentes disulfuro.

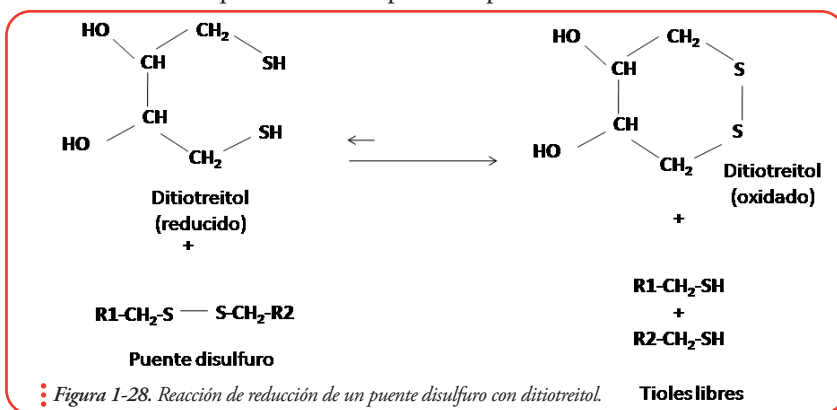


Figura 1-28. Reacción de reducción de un puente disulfuro con ditioneitol.

En la **Figura 1-28**. La reacción es muy eficiente. El ciclado del DTT conlleva la formación de un anillo químicamente muy estable de 6 miembros.

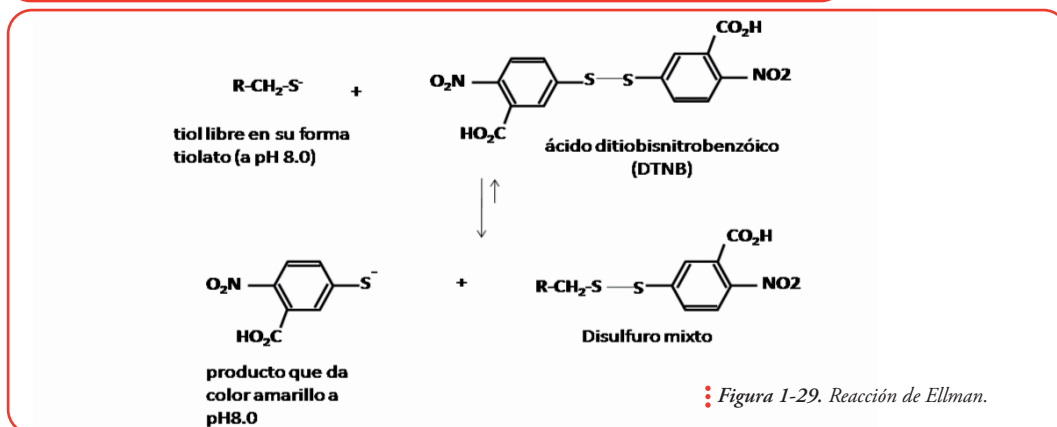


Figura 1-29. Reacción de Ellman.

Bases biofísicas de la estabilidad de la estructura proteica. Interacciones no covalentes

Como consecuencia de la adquisición de la estructura nativa, quedan determinadas orientaciones y distancias entre numerosos grupos funcionales cruciales para la función biológica. También queda determinada la capacidad que tienen las proteínas de interactuar con el medio.

¿Cómo se estabiliza la estructura proteica?, ¿Qué tipo de fuerzas están involucradas en el proceso de adquisición de la estructura proteica?, ¿Cómo interactúan las proteínas con el agua?

De manera general, se pueden reconocer distintos tipos de interacciones intramoleculares e intermoleculares que contribuyen con la estabilización o, por el contrario, con la desestabilización de la estructura proteica. Hoy sabemos que la estructura proteica resulta de la sumatoria de miles de interacciones.

Si bien la estructura molecular en término del tipo de enlace químico, que tiene lugar entre los átomos conectados en forma covalente, es crucial, hoy sabemos que la estructura proteica es consecuencia, por lo menos, en gran medida de un conjunto mayor de interacciones: las interacciones no covalentes. Vale la pena subrayar que todas las fuerzas entre átomos y moléculas son en origen electrostáticas, sin embargo, se pueden reconocer tipos bien definidos de interacciones:

- Interacciones iónicas y dipolares,
- Interacciones de van der Waals, enlaces por puente de hidrógeno, e
- Interacciones hidrofóbicas (ver efecto hidrofóbico).

Si bien las interacciones no covalentes aportan poco individualmente, en conjunto logran estabilizar, suficientemente, a la estructura proteica.

La complejidad y el desorden intrínseco de los líquidos (del agua en nuestro caso) hacen que no sea sencillo evaluar las propiedades de las interacciones no covalentes en este tipo de sistemas. Para entender a través de modelos y simulaciones el efecto de las interacciones en la estructura proteica hay que recurrir, en algunos casos, a modelos complejos que intentan rescatar ciertas propiedades del solvente (modelos de solvente implícito) y, en algunos casos, directamente hay que modelar, explícitamente, al solvente y a las interacciones entre el solvente y la macromolécula que se desea investigar: esto es, concretamente, agregar moléculas de agua en una caja imaginaria con dimensiones límites muy particulares en la que introducimos el modelo de la macromolécula (**Figura 1-30**).

Adelantándonos al problema del plegado proteico es importante, de aquí en más, tener en cuenta que las características de la superficie proteica determinan en gran medida las propiedades termodinámicas de la macromolécula, por ejemplo la estabilidad termodinámica y la diferencia de capacidad calorífica entre el estado nativo y el desplegado (ver más adelante).

Debemos tener en cuenta que, en una proteína, podremos encontrar miles de interacciones de van der Waals y cientos de puentes de hidrógeno, cada puente de hidrógeno podrá aportar en promedio unas 2 kcal/mol y cada interacción de van der Waals lo hará en 1 kcal/mol, las interacciones iónicas entre 1 y 3 kcal/mol.

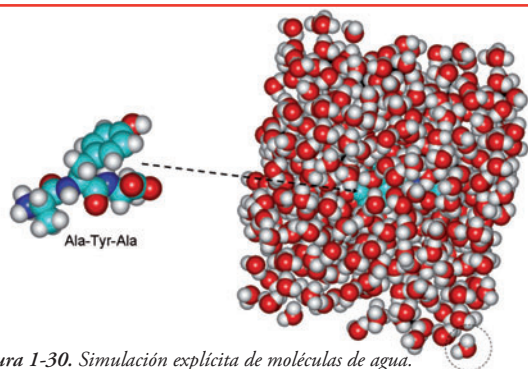


Figura 1-30. Simulación explícita de moléculas de agua.

En la Figura 1-30 se muestra el tripéptido Ala-Tyr-Ala y el mismo oligopéptido rodeado por unas 540 moléculas de agua que podrán interactuar entre sí y con el oligopéptido. Una de las 540 moléculas se resalta con un círculo en línea discontinua.

Interacciones electrostáticas

Como ya se ha mencionado con anterioridad, todas las fuerzas intramoleculares son, en principio, esencialmente electrostáticas. La ley de Coulomb dice que para dos átomos (por ejemplo A y B) en el vacío la energía de la interacción electrostática es, simplemente, el producto de sus cargas dividido por la distancia entre ellos (en este caso la distancia es r_{AB}):

$$\Delta E = \frac{Z_A \times Z_B \times \epsilon^2}{r_{AB}}$$

Aquí, ϵ es la carga de un electrón y Z es el número de cargas equivalentes de cada grupo. Si las dos cargas son de signo opuesto, entonces la energía disminuye a medida que ambos grupos se acercan. Así, la interacción resulta favorable. Por el contrario, si ambas cargas son del mismo signo, existirá repulsión entre ambos grupos. Es importante notar que tal cual está escrita esta ley vale, únicamente, en el vacío y para cargas puntuales. Este no suele ser el caso de las proteínas. En nuestro caso la constante dieléctrica del medio afectará, ampliamente, este tipo de interacciones. Por este motivo el entorno químico de las cargas definirá la potencia de la interacción. Si suponemos un ambiente, suficientemente, homogéneo entonces podemos aplicar una constante dieléctrica única y la ecuación de arriba se transforma en:

$$\Delta E = \frac{Z_A \times Z_B \times \epsilon^2}{D r_{AB}}$$

D es la constante dieléctrica. En el caso del agua, D es cercana a 80 (y dentro de una proteína puede tomar valores cercanos a 2-4). Se puede ver cómo disminuye la atracción de dos cargas de signo opuesto en un medio acuoso a medida que aumenta la constante dieléctrica ya que el valor de constante dieléctrica está en el denominador. Se muestran algunos valores de constante dieléctrica; Vacío, 1; Aire (seco), 1,00059; Cloroformo, 4,81; Metanol, 30; Agua, 80.

Para comprender el efecto del solvente sobre la carga se puede suponer, como se ve en la Figura 1-31, la existencia de una carga negativa ($-q$) que está rodeada por un gran volumen constituido por un dieléctrico (en nuestro caso el solvente y, en general, un dieléctrico es un material que no conduce la electricidad eficientemente, si la conduce muy ineficientemente,

entonces puede ser utilizado como aislante eléctrico). La carga $-q$ polariza el dieléctrico al punto que se forman dipolos cuya densidad de carga positiva apunta hacia la carga $-q$. Las cargas positivas neutralizarán parcialmente la carga $-q$; así, la carga efectiva ya no será $-q$ sino

$$-(q - \Sigma\delta q) = \frac{-q}{D}$$

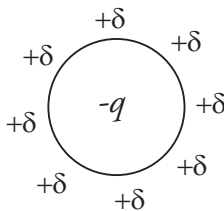


Figura 1-31. Carga efectiva.
La carga $-q$ polariza el dieléctrico.

Sin embargo en el caso de las proteínas no podemos considerar un entorno homogéneo: el agua del solvente, la superficie de la proteína con su esfera de hidratación y el corazón hidrofóbico de la proteína plegada poseen propiedades dieléctricas bien diferentes. Para este tipo de sistemas, altamente complejos, se debe considerar de manera explícita el solvente. Por esto debe calcularse la interacción de la macromolécula inmersa en un conjunto de moléculas de agua que deben modelarse, como ya se dijo explícitamente (una por una). A esto puede añadirse algunos condimentos como sales (por ejemplo cloruro de sodio, NaCl) que también alteran la constante dieléctrica del medio.

El pH y las proteínas

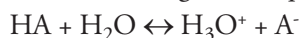
El pH es una medida de la acidez (o basicidad) de una solución. El pH es la concentración de cationes hidrógeno $[H^+]$ (protones) presentes en una solución. Así fue definido por el químico danés Sørensen, quien lo definió como el logaritmo negativo de base 10 de la actividad de los iones hidrógeno (de la concentración en soluciones diluídas). La definición que parece muy arbitraria, finalmente, resulta muy práctica.

$$pH = -\log [H^+]$$

Por ejemplo, el agua tiene un $pH=7,0$, la concentración de protones es $[H^+] = 1 \times 10^{-7}$ M (0,0000001 M) ya que $-\log[10^{-7}] = 7,0$; $pH = 7,0$ indica la neutralidad de la solución, la concentración de $[OH^-]$ en la solución deberá ser también 0,0000001 M para garantizar la neutralidad.

Son ácidas las soluciones con pH menores a 7, y básicas las que tienen pH mayores a 7. A pH 7,0 los grupos amino y carboxilo están cargados pero en el rango que va desde pH 1,0 a pH 14,0 estos grupos químicos presentan una serie de equilibrios en los que está implicada la unión y la disociación de un protón. Comprender el comportamiento ácido-base de los aminoácidos es importante porque éste, claramente, tiene efectos determinantes sobre la estructura y en la función de las proteínas.

Estos grupos se ionizan de acuerdo con el siguiente equilibrio:



Donde HA es el grupo donador de protón (en las proteínas $-\text{COOH}$, NH_3^+ o $-\text{SH}$) y A^- es el aceptor de protón ($-\text{COO}^-$, $-\text{NH}_2$ o $-\text{S}^-$)

Planteado el equilibrio se puede calcular la constante de equilibrio

$$K = \frac{[\text{H}^+][\text{A}^-]}{\text{HA}}$$

Se podrá derivar entonces

$$[\text{H}^+] = \frac{K \times \text{HA}}{[\text{A}^-]}$$

$$-\log [\text{H}^+] = -\log \left(\frac{K \times \text{HA}}{[\text{A}^-]} \right)$$

$$-\log [\text{H}^+] = -\log \left(K \times \frac{\text{HA}}{\text{A}^-} \right)$$

$$-\log [\text{H}^+] = -\log_{10} K - \log_{10} \frac{\text{HA}}{\text{A}^-}$$

$$-\log_{10} [\text{H}^+] = -\log_{10} K + \log_{10} \frac{\text{A}^-}{\text{HA}}$$

Por último usando la definición de pH

$$\text{pH} = \text{p}K + \log_{10} \frac{\text{A}^-}{\text{HA}}$$

Como ejemplo, en la **Figura 1-32** se muestra el equilibrio de ionización para un residuo de aminoácido (serina). A pHs muy ácidos ($\text{pH} < 2$) la especie mayoritaria será la cargada positivamente a pH 2,2 coexistirán (concentraciones idénticas) la especie cargada, positivamente, con la especie neutra (una carga positiva más una negativa). Se puede hacer la cuenta para evaluar esta igualdad de concentraciones:

$$2,2 = 2,2 + \log \frac{\text{A}^-}{\text{HA}}$$

$$\log 1 = \log \frac{\text{A}^-}{\text{HA}}$$

$$\text{HA} = \text{A}^-$$

Elevar aún más el pH de la solución conducirá a la pérdida de un protón del grupo amino y a la formación de especies conteniendo una carga negativa pasando por un segundo $\text{p}K$ ($\text{p}K=9,2$). Otra vez, en este pH ($\text{pH}= 9,2$) las concentraciones de la especie neutra y de la especie cargada negativamente serán iguales.

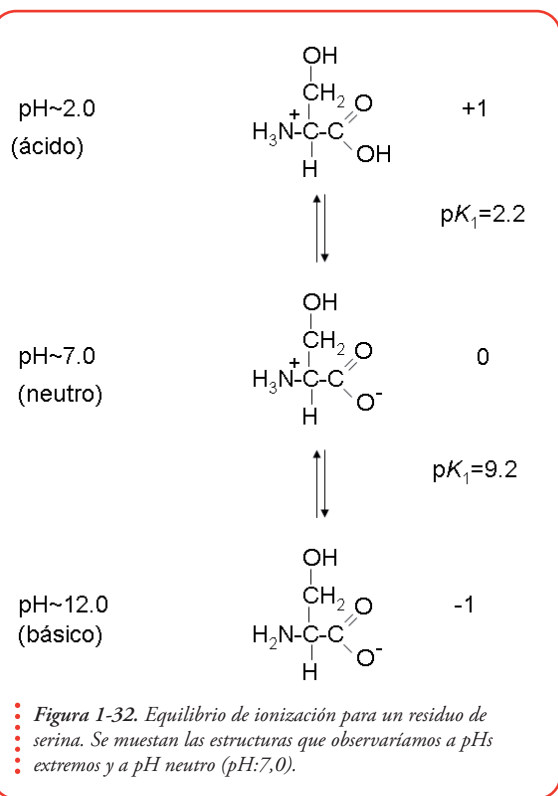
Aquí puede verse que a pH neutro la especie predominante será la *zwiteriónica* que, si bien es altamente bipolar, su carga neta es 0 y esto determinará que no migre en un campo

eléctrico hacia ninguno de los dos polos. El pH en el que la carga neta de un aminoácido, de un péptido o de una proteína, es cero se conoce como punto isoeléctrico o pI. Para los aminoácidos que contienen grupos que pueden funcionar como donores o aceptores de protones se observarán equilibrios más complejos; los residuos que muestran valores de pK adicionales debido a grupos funcionales en la cadena lateral son: lisina, arginina, histidina, aspartato, glutamato, cisteína y tiro-sina. En la **Tabla 1-2** se muestran los valores de pK para estas residuos.

	pK ₁	pK ₂	pK _R
Arg	1,8	9,0	12,5
Lys	2,2	9,1	10,5
His	1,8	9,3	6,0
Asp	2,0	9,9	3,9
Glu	2,1	9,5	4,1
Cys	1,9	10,7	8,4
Tyr	2,2	9,2	10,5

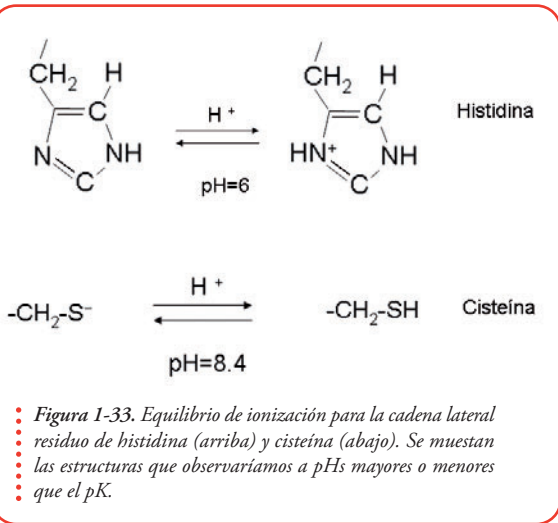
• **Tabla 1-2.** Constantes de equilibrio adicionales debido a grupos funcionales en la cadena lateral. Valores de pK para residuos con cadenas laterales (R) que presentan grupos químicos que pueden funcionar como donores o aceptores de protones y por lo tanto presentan un pK_R.

Así por ejemplo el residuo de histidina (H) is) tiene un pK para la cadena lateral de 6,0, esto se traduce en que a pH 7,5 las histidinas serán neutras pero a pH 6,0 cerca del 50 % de las histidinas estarán protonadas y se habrán cargado (Figura 1-33, panel superior). Por otro lado, el grupo tiol (-SH) de las cisteínas (Cys) podrá estar en equilibrio con la forma tiolato (cargada negativamente); a pH 8,4 el 50 % de las cisteínas libres (aquellas que no forman puentes disulfuro, ver más adelante) debería estar en su forma tiolato (-S⁻ **Figura 1-33**, panel inferior). Sin embargo la estructura primaria (residuos vecinos en secuencia) y el plegado (al posicionar residuos vecinos en el espacio) de las proteínas puede modular increíblemente el valor de pK para estos equilibrios.



• **Figura 1-32.** Equilibrio de ionización para un residuo de serina. Se muestran las estructuras que observaríamos a pHs extremos y a pH neutro (pH:7,0).

35



• **Figura 1-33.** Equilibrio de ionización para la cadena lateral residuo de histidina (arriba) y cisteína (abajo). Se muestran las estructuras que observaríamos a pHs mayores o menores que el pK.

Por ejemplo en algunos casos se observan cisteínas que poseen un pK-5.

Interacciones de van der Waals

Curiosamente y, aunque suene un poco paradójico, una de las interacciones más importantes, energética y estructuralmente, es la repulsión que tiene lugar entre dos grupos cuando estos se acercan más allá de una distancia límite. Los átomos y/o las moléculas no pueden aproximarse arbitrariamente entre sí. Dicho de otro modo, dos grupos no pueden ocupar simultáneamente la misma región del espacio. Esto tiene una base en el principio de exclusión de Pauli (un principio cuántico que establece que no puede haber dos *fermiones* (los protones, los neutrones, y los electrones son fermiones, pero los fotones por ejemplo no lo son) con todos sus números cuánticos idénticos. Concretamente, existen restricciones Geométricas impuestas por los orbitales moleculares y repulsión electrostática causada por los electrones externos.

Como la energía de repulsión crece muy abruptamente más allá de la distancia límite (ver interacción de van der Waals óptima) se puede considerar que los átomos y moléculas que interactúan tienen dimensiones definidas y ocupan volúmenes que son impenetrables por otros átomos. Así, podemos modelar a los átomos como esferas, y el volumen impenetrable se define como volumen de van der Waals.

El volumen de la proteína entera puede modelarse como la suma de la contribución de cada residuo de aminoácido. Los radios de van der Waals (Tabla 1-3) definen también la superficie de van der Waals. Esta superficie estará caracterizada por la presencia de numerosas grietas y rugosidades. Podemos ver a partir de este tipo de modelo, cómo una proteína comienza a tomar la apariencia de objeto.

En la Figura 1-35 se muestra un esquema en el que se comparan dos distancias entre átomos de carbono: dos carbonos unidos covalentemente (1,5 Å) y dos carbonos a distancia de van der Waals (3,4 Å).

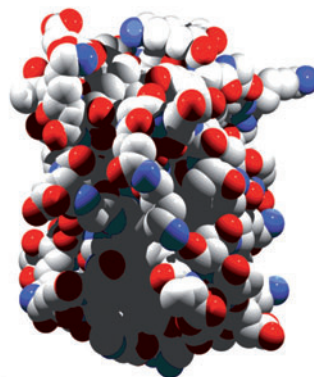


Figura 1-34. Representación de los volúmenes de van der Waals de los átomos de C (gris), N (azul), O (rojo) y S (amarillo) de la tiorredoxina de *E. coli*.

Radios de van der Waals

Tipo de átomo	rvdw (nm)
H (aromático)	0,10
H (alifático)	0,12
O	0,15
N	0,16
C	0,17
S	0,18

Tabla 1-3. Radios de van der Waals. Los radios de van der Waals (rvdw) medidos en nm.

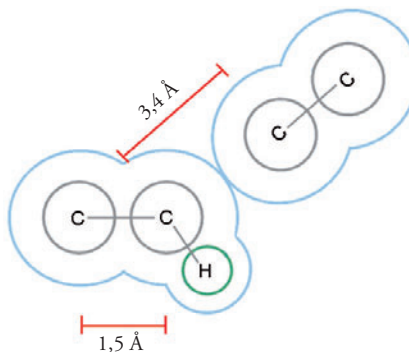
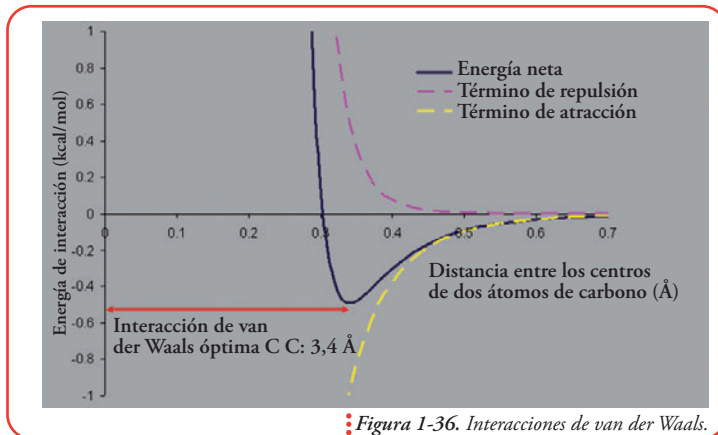


Figura 1-35. Enlace covalentes e interacciones de van der Waals. Dos carbonos unidos covalentemente (1,5 Å) y dos carbonos a distancia de van der Waals (3,4 Å).

Los átomos que no están unidos, covalentemente, pueden atraerse unos a otros, inclusive en ausencia de grupos cargados. Esto ocurre como resultado de interacciones mutuas, en muchos casos, producto de procesos de inducción de polarización (inducción de densidades de carga). Este tipo de interacción atractiva débil se conoce como interacción de Van der Waals; puede provenir de la interacción de dos dipolos permanentes, alternativamente de la interacción entre un dipolo permanente y un dipolo inducido e incluso de la interacción entre dos dipolos, mutuamente, inducidos.

Las interacciones de Van der Waals suelen representarse como la energía potencial en función de la distancia (d) que incluye tanto fuerzas atractivas como repulsivas, se trata de un balance entre ambas fuerzas (Figura 1-36).



En la Figura 1-36 se grafica la energía en kcal/mol para la interacción entre dos átomos de carbono. Existe una distancia óptima de 3,4 Å en la que la energía es mínima. Se usó un modelo de potencial de tipo Lennard-Jones (6-12).

La siguiente función describe este proceso:

$$V_{\text{vdw}} = \frac{A}{r^m} - \frac{B}{r^6}$$

Aquí, V_{vdw} es el potencial (energía potencial) de van der Waals. A y B son constantes que describen la magnitud de los términos atractivo y repulsivo. La función de energía potencial puede ser modelada por el potencial de Lennard-Jones, en este caso el exponente m toma el valor de 12 (la función se conoce entonces como 6-12). Cuando r , la distancia entre ambos átomos, es $r = \sum r_{\text{vdw}} \Rightarrow r = r_0$, es decir la suma de los radios de van der Waals, la función define una distancia óptima, de mínima energía. Sin embargo, a distancias más cortas se produce un abrupto aumento en la energía (V_{vdw}), la parte repulsiva se representa con un exponente $m = 12$. En la Tabla 1-4 se muestran coeficientes para los términos repulsivos (A) y atractivos (B) de potenciales van der Waals (tipo 6-12, Lennard-Jones). También se muestran los valores de r_0 correspondientes a la interacción atómica.

$$V(r) = \frac{1,18 \times 10^{-6}}{r^{12}} - \frac{1,52 \times 10^{-3}}{r^6}$$

Vale la pena, como ejercicio construir una planilla de cálculo con un r variable (por ejemplo entre 0.1 y 1) y nuestra variable dependiente, la energía potencial, será por ejemplo para dos átomos de carbono (C...C) no unidos covalentemente $V(r)$. En este caso nuestro r_0 , el r en el que se hace mínimo $V(r)$ será 0.34 nm.



Interacción atómica	A (kJ nm ¹² /mol)	B (kJ nm ⁶ /mol)	r ₀ (nm)
H...H	1,84 × 10 ⁻⁸	1,92 × 10 ⁻⁴	0,240
H...C	1,57 × 10 ⁻⁷	5,27 × 10 ⁻⁴	0,290
H...N	1,11 × 10 ⁻⁷	5,15 × 10 ⁻⁴	0,275
H...O	1,03 × 10 ⁻⁷	5,11 × 10 ⁻⁴	0,272
C...N	8,90 × 10 ⁻⁷	1,51 × 10 ⁻³	0,325
C...O	8,49 × 10 ⁻⁷	1,51 × 10 ⁻³	0,322
C...C	1,18 × 10 ⁻⁶	1,52 × 10 ⁻³	0,340
N...N	6,63 × 10 ⁻⁷	1,50 × 10 ⁻³	0,310
N...O	6,30 × 10 ⁻⁷	1,50 × 10 ⁻³	0,307
O...O	5,97 × 10 ⁻⁷	1,51 × 10 ⁻³	0,304

Tabla 1-4. Coeficientes para los términos repulsivos (A) y atractivos (B) de potenciales van der Waals (tipo 6-12, Lennard-Jones). La distancia entre los centros de ambos átomos (r₀) se adjunta en nm.

Unión por puente de hidrógeno

Un tipo de interacción, especialmente, importante en sistemas biológicos y, particularmente, en lo relacionado a la estructura proteica es la unión por puente de hidrógeno.

Como ya se ha mencionado, Linus Pauling fue el primer investigador que realizó los primeros cálculos para proteínas sobre la base de modelos experimentales de moléculas pequeñas postulando la importancia de los puentes de hidrógeno en el proceso de la adquisición de estructura (plegado proteico) y en su estabilización.

Pero, ¿qué es un puente de hidrógeno?

Es una interacción entre un **enlace polarizable**, un átomo de hidrógeno unido a un átomo al que llamaremos **donor** o **D**, y orbitales también polarizables de un tercer átomo al que llamaremos **aceptor**, **A**. En cierto sentido se parece a una interacción no covalente del tipo diplo-dipolo. El principal componente de un puente de hidrógeno es una interacción electrostática: el átomo de hidrógeno tiene densidad de carga positiva y carga parcial negativa sobre los átomos electronegativos.



En un puente de hidrógeno, dos átomos electronegativos compiten fuertemente por el mismo átomo de hidrógeno. *Según Linus Pauling, la electronegatividad es la capacidad de un átomo para atraer hacia sí electrones de otro átomo.* Esta propiedad depende de su estado de oxidación. Así, un mismo elemento puede presentar distintas electronegatividades dependiendo del tipo de molécula o entorno químico en el que se encuentre.

Es interesante observar que la distancia óptima que separa donador y aceptor en ciertos puentes es menor que la suma de los respectivos radios de van der Waals de los átomos involu-

mecánico-cuántico de los orbitales moleculares (las regiones del espacio con probabilidad de ser visitadas por electrones). En particular, aquellos átomos que sufran modificaciones en su estructura electrónica deben ser considerados especialmente.

A las funciones que describen la energía del sistema (de la conformación) las llamamos campos de fuerza. Los campos de fuerza deben estar calibrados con datos experimentales.

$$E_{total} = \sum_{\text{enlaces}} K_r (r - r_{eq})^2 + \sum_{\text{ángulos}} K_\theta (\theta - \theta_{eq})^2 + \sum_{\text{driedos}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i \neq j} \frac{A_{ij}}{R^{12}_{ij}} - \frac{B_{ij}}{R^6_{ij}} + \frac{q_i q_j}{DR_{ij}}$$

Un ejemplo muy simplificado de campo de fuerza se muestra en la ecuación anterior.

En este campo una parte de la ecuación intenta describir la energía relacionada a perturbaciones del enlace covalente: si el enlace tiene una distancia óptima de menor energía a la que llamaremos distancia de equilibrio (r_{eq}) una perturbación del tipo de estiramiento o contracción (por ejemplo $r > r_{eq}$) incrementará la energía del sistema (el sistema sale de la situación de equilibrio, aumenta su energía). K_r es una constante de proporcionalidad. Este tipo de potencial es similar al que se utiliza para describir el estiramiento de un resorte y aproxima la contribución de cada enlace a energía potencial como una parábola. Por otro lado, se incluye un término que representa el efecto de perturbaciones sobre cada ángulo (es una sumatoria que incluye a todos los ángulos entre tres átomos conectados entre sí por enlaces covalentes en nuestro sistema) de equilibrio (θ_{eq}) formado por 3 átomos (dos enlaces). Un incremento o una disminución de este ángulo apartarían al sistema de la situación de equilibrio. En este caso también se simula el potencial con una función que describe una parábola (cuyo mínimo es el ángulo θ_{eq}). Este potencial también considera, explícitamente, los ángulos diedros (ver estructura secundaria) y cómo se apartan del ángulo de mínima energía (ϕ). El último término incluye energías de van der Waals e interacciones electrostáticas. Para realizar el cálculo cada átomo y cada enlace debe, previamente, ser extensivamente caracterizado

Caracterizado o parametrizado: carga neta o carga parcial, volumen de van der Waals, constantes de estiramiento de los enlaces covalentes, ángulos diedros óptimos y otras características en el caso de que se utilice un campo de fuerza más complejo. Por ejemplo, los campos de fuerza comúnmente usados contienen términos que incluyen explícitamente un componente de energía relacionado con la formación de puentes de hidrógeno, y otros componentes que incluyen la solvatación. En estos casos interesará contar con una buena parametrización de la capacidad de formar puentes de hidrógeno de cada átomo, y del incremento de energía por transferencia a entorno acuoso.

Efecto hidrofóbico

Como el agua forma una red de puentes de hidrógeno, la disolución de un compuesto apolar (hidrofóbico) causará un rearrreglo local de esta red de puentes. En el entorno del compuesto apolar las moléculas de agua se reordenan y se alinean. Así el número de puentes de hidrógeno que pueden establecer entre sí se maximiza (no pueden establecer este tipo de interacción con el compuesto apolar). Sin embargo el aumento del orden (disminuye la entropía del solvente) como consecuencia del alineamiento de las moléculas de solvente en las

inmediaciones del compuesto tiene un costo alto. Es importante ver que dos moléculas con superficie apolar, entonces, tenderán a establecer contactos intermoleculares de estas superficies, minimizando la superficie apolar accesible al solvente y, por ende, minimizando el reordenamiento de moléculas de agua que es entrópicamente desfavorable. En el caso de las proteínas, cada residuo de aminoácido tiene una hidrofobicidad particular y una manera de medirla; es justamente medir cómo se *particionan* (cómo se distribuyen) las moléculas entre dos fases: una fase apolar, típicamente un solvente orgánico (ciclohexano por ejemplo) y otra fase polar (H_2O). La distribución tendrá que ver con cómo varía la energía del sistema en función de esta distribución.

En la tabla se muestran las energías libres de la transferencia de cadenas laterales de aminoácidos desde ciclohexano a H_2O (energía libre (ΔG) es un parámetro termodinámico sumamente importante porque indica, directamente, si un proceso ocurrirá espontáneamente o no: si el valor es negativo ($\Delta G < 0$), el proceso es espontáneo y no lo será si es positivo ($\Delta G > 0$) y deberá estar acoplado a otro que sí lo sea; sin embargo no nos da cuenta sobre la velocidad con la que ocurrirá dicho proceso). Se puede ver muy directamente cómo transferir un residuo apolar desde ciclohexano a H_2O es energéticamente costoso. También es muy interesante, ya a esta altura, tener en cuenta que podríamos agregar a la solución acuosa distinto tipo de compuestos que alteren las propiedades de solubilidad de los aminoácidos y que modifiquen la energía de estos en solución acuosa.

Existe una gama de compuestos llamados caótropos que aumentan la solubilidad de cadenas laterales de aminoácidos apolares. Así, por ejemplo el costo de mantener en solución acuosa leucina o valina, no es tan elevado. Entre estos compuestos los más comunes son la urea y el cloruro de guanidinio (**Figura 1-37**).

Este tipo de compuestos bajarán mucho la energía de las proteínas en estado desplegado, favoreciendo su estabilización con respecto a otros estados conformacionales en solución. Como se acaba remencionar, esto es posible porque dichos compuestos solubilizan las cadenas laterales apolares que formaban parte del corazón de la proteína.

Por otro lado, existen otros tipos de compuestos a los que llamaremos, en general, os-

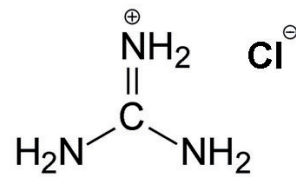
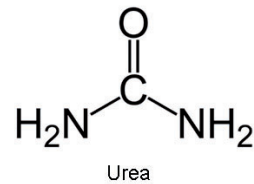


Figura 1-37. Estructura química de dos caótropos comúnmente usados.

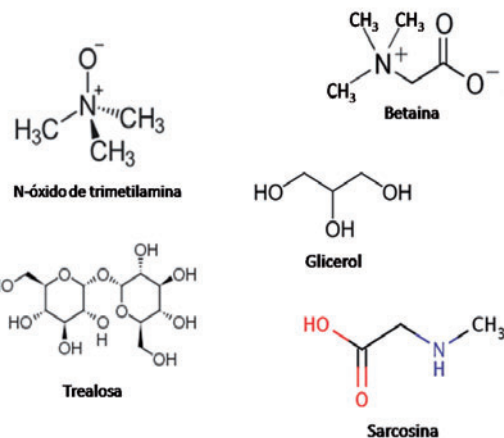


Figura 1-38. Algunos de los osmolitos utilizados más comúnmente.

molitos, que actúan exactamente al revés: bajan la solubilidad de los aminoácidos apolares. Por esto, los estados desplegados, que exponen residuos apolares al solvente, se desestabilizarán con respecto a los estados más compactos, con menor superficie expuesta. Algunos de estos compuestos, entre muchos otros, son el sulfato de sodio, el glicerol, la trealosa, el N-óxido de trimetilamina (Figura 1-38).

Estructura Terciaria, dominios, topología.

En ingeniería, los sistemas con cierta complejidad, suelen construirse combinando partes o módulos. Cada uno de estos módulos tiene relativa independencia y el sistema suele depender, por lo menos en parte, de cada una de ellos. Por otro lado, existen propiedades que son consecuencia del arreglo de las partes. Son propiedades del sistema y no de algún módulo en particular. Los módulos pueden ser extraídos de un sistema y usados como parte en otro. Para esto tiene que existir un mínimo de compatibilidad entre sistemas. Además, el agregado de nuevos módulos o la exclusión de partes del sistema suele ser complicado porque, también, debe satisfacer ciertos requisitos de compatibilidad con los módulos preexistentes.

Podemos ver a las proteínas como un arreglo de partes, un arreglo de elementos de estructura secundaria. Este arreglo estructural define en el caso de las proteínas una unidad de plegado independiente o una unidad autónoma de plegado: un dominio.

Las interacciones existentes entre elementos de estructura secundaria (en conjunto con las interacciones locales dentro de cada elemento) permiten la estabilización de motivos y del dominio estructural, la estabilización de lo que conocemos con el nombre de estructura terciaria. Concretamente, un dominio es la unidad fundamental de estructura terciaria. A estas interacciones, que no son locales (no son entre dos residuos pertenecientes al mismo elemento de estructura secundaria) las llamamos interacciones terciarias. El empaquetamiento fino de las cadenas laterales de la hélice C-terminal de la proteína tiorredoxina con el resto de la macromolécula, (mediado por interacciones de van der Waals, entre otras) es un excelente ejemplo de interacciones terciarias.

En la **Figura 1-39** se muestran ejemplos de interacciones terciarias formadas en el entorno de la leucina 103 de la tiorredoxina. También se muestran algunas interacciones terciarias entre residuos cargados o polares presentes en la superficie de la macromolécula.

En muchos casos podremos eliminar alguna de las partes sin afectar, significativamente, la integridad de la estructura proteica, simplemente desestabilizaremos a la es-

Cadena lateral	Ciclo hexano→H ₂ O kcal/mol
Leu	4,92
Ile	4,92
Val	4,04
Pro	3,58
Phe	2,98
Met	2,35
Trp	2,33
Ala	1,81
Cys	1,28
Gly	0,94
Tyr	-0,14
Thr	-2,57
Ser	-3,40
His	-4,66
Gln	-5,54
Lys	-5,55
Asn	-6,64
Glu	-6,81
Asp	-8,72
Arg	-14,92

••• *Tabla 1-5. Valores de energía libre para la transferencia de aminoácidos desde una fase apolar de ciclo hexano a H₂O.*

estructura nativa. Por ejemplo, podemos reemplazar un residuo de leucina (Leu) por alanina (Ala) acortando la cadena lateral y generando en el interior de la proteína una cavidad (un espacio vacío).

Inclusive, a veces, podemos eliminar por completo un segmento de la cadena polipeptídica sin comprometer, por completo, la estabilidad de la proteína. Debemos tener en cuenta que, en numerosos casos, la eliminación de una parte puede traer consecuencias drásticas sobre la estabilidad de la estructura terciaria y, por lo tanto, la proteína puede no plegarse, plegarse parcialmente o alternativamente plegarse de manera incorrecta.

La estabilidad de la estructura secundaria de cada uno de los elementos aislados puede ser muy variable. Para comprender el significado del término *elemento aislado* se puede ver el esquema del péptido 94-108 de unos

quince aminoácidos que corresponde a una de las hélices α en la estructura correctamente plegada de la tiorredoxina; este péptido se puede sintetizar mediante una serie de reacciones químicas complejas, luego puede ser separado del resto de los componentes de la mezcla de reacción mediante un protocolo de purificación del péptido. Una vez purificado podemos hacer experimentos para determinar el contenido de estructura secundaria de este pequeño fragmento de la proteína tiorredoxina.

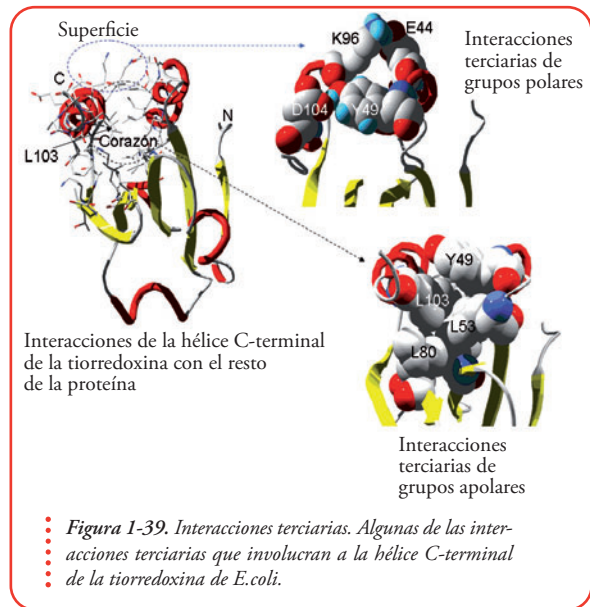
En algunos casos (como en el ejemplo mencionado inmediatamente arriba) los elementos pueden requerir interacciones terciarias (con otros elementos) que contribuyan, simultáneamente, con la estabilidad de la estructura secundaria del péptido (en este caso una hélice α) y con la estabilidad de la estructura terciaria de la proteína.

Estos péptidos que requieren estabilización terciaria y que, cuando están aislados no poseen estructura secundaria estable en solución acuosa, se estructuran cuando interactúan con el resto de la macromolécula y contribuyen con la estabilidad global de la proteína. El resto de la macromolécula, a veces, puede mimetizarse mediante el uso de otras moléculas (por ejemplo detergentes). La interacción del péptido con estas moléculas simula la interacción del péptido con la superficie del resto de la proteína.

Sin embargo, ciertos péptidos en ausencia del resto de la macromolécula pueden tener estructura secundaria estable.

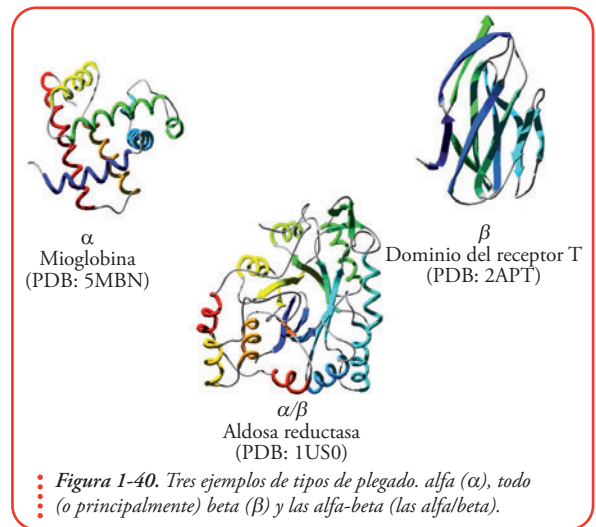
Más allá de que esto ocurra o no, la interacción entre elementos de estructura secundaria definirá de manera fina la topología de la proteína: el recorrido de la cadena polipeptídica en el espacio. Ha habido numerosos intentos por clasificar a los distintos dominios de las proteínas de acuerdo a su estructura secundaria y a su topología.

Según el contenido de estructura secundaria se pueden reconocer clases de plegado del tipo todo (o principalmente) alfa (α), todo (o principalmente) beta (β) y las alfa-beta (Figura 1-40). Este último conjunto o clase incluye tanto a las estructuras alfa+beta como a las



alfa/beta. Las estructuras alfa+beta corresponden a dominios en las que puede observarse un subdominio alfa en el que elementos helicoidales interactúan entre sí segregados de un subdominio beta, formado por hebras; en cambio los dominios alfa/beta tienen elementos alfa alternados con elementos beta. También podríamos agregar una cuarta clase de dominios que contienen muy bajo contenido de estructura secundaria.

Además podremos clasificar a los dominios, por la orientación y el aspecto que estos elementos toman en el espacio entre sí, por ejemplo si adoptan forma de barril, sándwich (por ejemplo $\alpha\beta\alpha$ o $\beta\alpha\beta$, paquete de hélices).



Familias de proteínas y proteínas multi dominio

Existen claras restricciones respecto a la posibilidad de empaquetamiento de los elementos de estructura secundaria para obtener el empaquetamiento óptimo que permita mantener los residuos hidrofóbicos en el corazón de las proteínas, probablemente, ésta sea una de las causas de que existan en la naturaleza solamente unos pocos miles de dominios estructurales conocidos (el espacio de conformaciones posibles de proteínas en la naturaleza parece ser acotado).

Dominios diferentes pero con el mismo tipo de plegado, es decir, con el mismo arreglo general de elementos de estructura secundaria y la misma conectividad general de estos elementos, muchas veces pueden tener algún elemento de estructura secundaria periférico o alguna conexión diferente, por ejemplo, pueden diferir en tamaño (en longitud en aminoácidos de la cadena en estas regiones). A pesar de estas diferencias podemos reconocer la topología general, el tipo de plegado (el *fold*). En los casos más extremos esas regiones diferentes pueden formar parte de un porcentaje importante del dominio y hacer la tarea de reconocimiento de tipo de plegado más complicada.

Aunque es difícil determinar el número de tipos de plegado diferentes conocidos (*fold*s), y este número varía según las definiciones y la fuente de donde se tome este dato, basándose en SCOP, una base de datos para la clasificación estructural de proteína para investigación creada por Alexey G. Murzin, Steven E. Brenner, Tim Hubbard y Cyrus Chothia, existirían cerca de **1.300 tipos de plegado distintos** (exactamente 1.283 tipos de plegado se han descrito hasta la fecha, 2009).

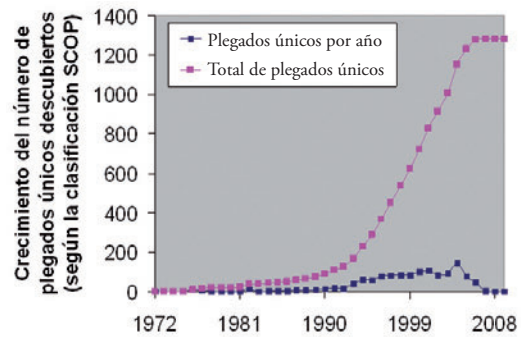
Una proteína puede estar formada por numerosos dominios. Todos ellos codificados en

Vale la pena recordar que, dentro de una proteína, un dominio es un arreglo de elementos de estructura secundaria que se estabiliza por sí mismo y, generalmente, se pliega independientemente del resto de la proteína (es una unidad autónoma de plegado).

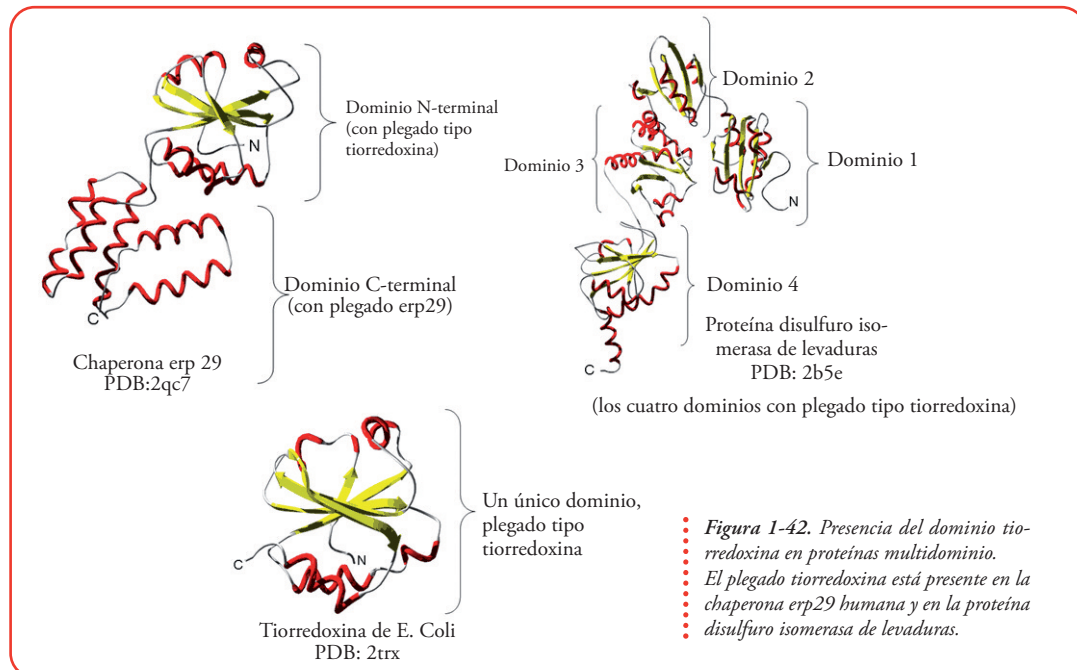
una única cadena polipeptídica. En muchos casos un dominio determinado aparece formando parte de una gran variedad de proteínas, cada una de estas proteínas puede tener así funciones comunes con el resto y otras funciones que son exclusivas.

Los dominios son frecuentemente nombrados según la función biológica pero, también, es común nombrarlos por el tipo de plegado. En la Figura 1-42 se muestra como ejemplo el tipo de plegado tiorredoxina. Este plegado está presente en muchas proteínas. Entre ellas, está presente en uno de los dos dominios de la proteína chaperona humana erp29. Esta proteína erp29 interviene en el proceso de plegado de otras proteínas en el retículo endoplasmático de la célula. En este caso el dominio tiorredoxina no tiene función tiorredoxina sino que, aparentemente, funciona como un módulo de dimerización interviniendo en el arreglo de estructura cuaternaria.

Es interesante ver que en los últimos dos años no se han descubierto nuevos tipos de plegado. Esto indicaría que el espacio de conformaciones proteicas existente en la biología podría ser ciertamente limitado (ver Figura 1-41).



••• Figura 1-41. El número de plegados distintos existentes en la naturaleza parece ser limitado. Estadística disponible en la Protein Data Bank.



El segundo ejemplo es una proteína disulfuro isomerasa de levaduras. Esta proteína tiene cuatro dominios tiorredoxina en tandem formando parte de la misma cadena polipeptídica. La función en este caso sí está más emparentada con la función de la tiorredoxina, esta proteína interviene en el rearreglo de puentes disulfuro de otras proteínas.

Familias y superfamilias (homólogos o análogos)

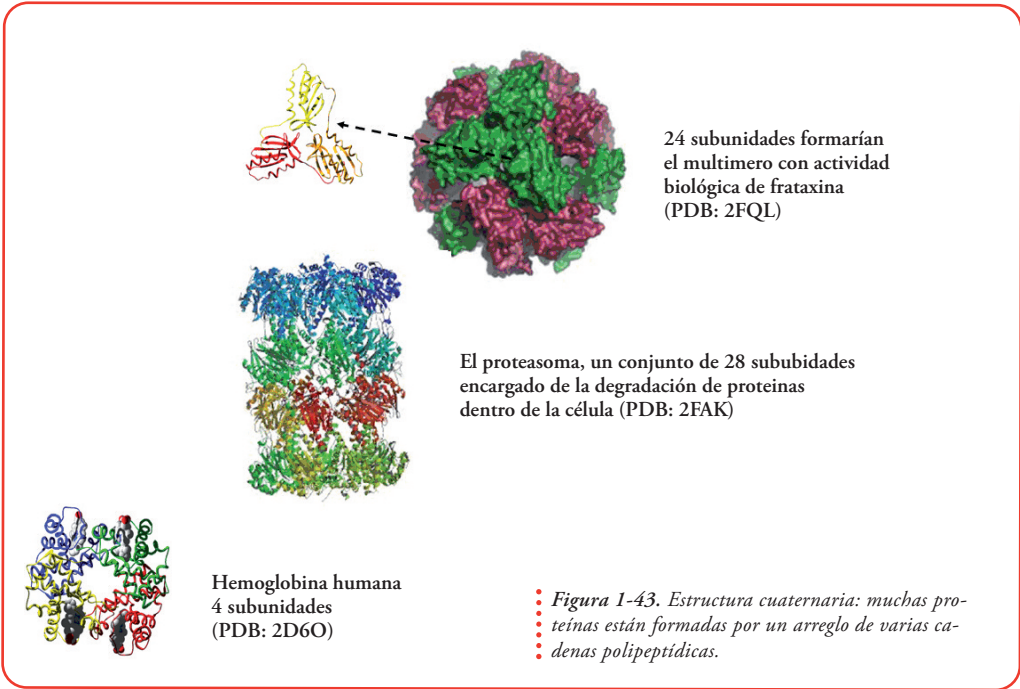
Durante el curso de la evolución, las proteínas han derivado de una proteína ancestro común por mutaciones, substituciones, deleciones, inserciones o recombinación, dando lugar a familias de proteínas homólogas. No todas las posiciones son igualmente susceptibles a cambios, en algunos casos es importante para la estructura o para la función mantener invariables algunas posiciones. Así, proteínas cuyas secuencias comparten identidad de secuencia > al 40 %, decimos que forman una familia y, probablemente, tengan la misma función. Proteínas homólogas más remotas, cuya identidad de secuencia es < al 30 %, se agrupan en una familia evolutiva más amplia o en una superfamilia. A veces la divergencia dentro de una familia es tan grande que es difícil reconocer proteínas homólogas si comparamos sólo la secuencia (las secuencias son muy diferentes entre sí). Homólogos remotos pueden reconocerse sólo comparando sus estructuras. Proteínas que no comparten un gen ancestro común y poseen el mismo arreglo tridimensional o plegamiento, son denominadas análogos.

Cuando no hay similitud de secuencia aparente y las proteínas muestran similitud estructural es difícil determinar si son dos proteínas extremadamente distantes de la misma familia evolutiva que divergieron en secuencia y función o si provienen de diferentes proteínas o dominios ancestros que convergieron en el mismo arreglo estructural o plegamiento, se elimina.

Estructura cuaternaria, subunidades

Muchas proteínas están formadas por un arreglo de varias subunidades, es decir por varias cadenas polipeptídicas (Figura 1-43). Estas subunidades pueden ser idénticas o pueden no serlo. En muchos casos la función o actividad biológica está asociada a residuos de la interfase entre subunidades.

Uno de nuestros ejemplos favoritos, la frataxina, es un caso en el que las subunidades que se asocian son idénticas (en este caso son 24 subunidades). Por ahora contamos, únicamente, con la estructura de uno de los componentes de este multímero, el trimero. Otro ejemplo, interesante es el proteasoma, una especie de máquina que se encarga de degradar proteínas convirtiéndolas en pequeños péptidos en el citoplasma celular. Por último un ejemplo clásico es la hemoglobina, compuesta por cuatro subunidades que unen oxígeno en forma cooperativa. Podríamos citar cientos de ejemplos, pero los lectores podrán encontrarlos en libros de texto y en las bases de datos de estructuras de proteínas.

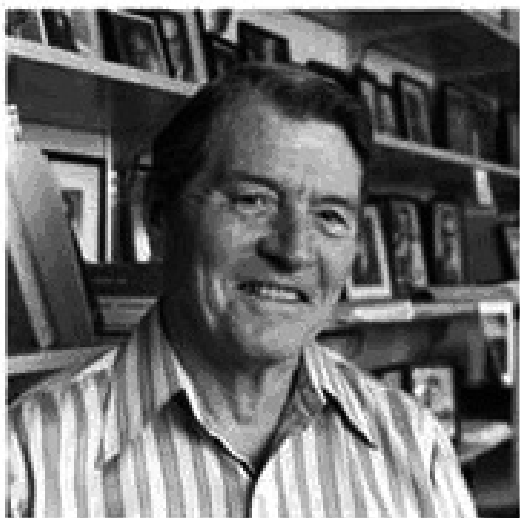


Arriba, la estructura trimérica de una variante de frataxina de levadura puede ajustarse en una imagen obtenida por microscopía electrónica de un complejo mayor (24 subunidades); en el centro, la estructura correspondiente al proteasoma, un arreglo 3 anillos de 7 subunidades cada uno); abajo la estructura de la hemoglobina humana formada por cuatro subunidades. (ver Figura 1-43.)



El Proceso de Plegado Proteico

Segunda Parte



Christian Anfinsen

El proceso de plegado es espontáneo:

El descubrimiento de Anfinsen y sus implicancias

Tal vez uno de los descubrimientos más importantes de todos los tiempos fue el descubrimiento realizado por Christian Anfinsen. Como ya se ha mencionado en el **Capítulo I**, sus experimentos condujeron a la enunciación de la hipótesis termodinámica del plegado proteico.

En pocas palabras, las proteínas se pliegan, espontáneamente, y el estado plegado corresponde al estado de mínima energía, es decir el más estable. Así podríamos, en principio, producir una proteína y ésta debería llegar a su forma correctamente, plegada (nativa) y lo que resulta realmente importante, es que lo hace en ausencia de otros factores biológicos. Condiciones adecuadas de solvente deberían ser suficientes para estabilizar dicha conformación.

Para poder entender esto es conveniente tener en cuenta, al menos, las principales características de los distintos estados conformacionales. También es fundamental no perder de vista que las distintas conformaciones están en equilibrio en la solución y que, algunas, en determinadas condiciones experimentales, pueden ser suficientemente inestables y poco probables; en definitiva como para que la solución no se pueble de ellas.

Para poder entender esto es conveniente tener en cuenta, al menos, las principales características de los distintos estados conformacionales. También es fundamental no perder de vista que las distintas conformaciones están en equilibrio en la solución y que, algunas, en determinadas condiciones experimentales, pueden ser suficientemente inestables y poco probables; en definitiva como para que la solución no se pueble de ellas.

El estado nativo

La compactación óptima de las cadenas laterales de los residuos es una de las características más notables del estado nativo; por otro lado, la mayor parte de los grupos polares se encuentran en la superficie, en el exterior, en contacto con el solvente. Además, y como consecuencia del alto grado de compactación, el estado nativo tiene mínima superficie expuesta al solvente, particularmente, mínima superficie apolar expuesta: las proteínas nativas tienen un corazón hidrofóbico.

Típicamente y salvo raras excepciones, el estado nativo es el estado conformacional que tiene actividad biológica, así podemos examinar el contenido de proteína en estado nativo

en una solución midiendo la actividad biológica de dicha proteína; típicamente si se tratara de una enzima podríamos medir la actividad enzimática.

Como ya hemos visto una de las características más importantes de la estructura nativa es que existen interacciones terciarias fijas que mantienen empaquetados a los elementos de estructura secundaria entre sí. Las cadenas laterales, entonces, se ubican en el interior de la proteína en entornos asimétricos desde el punto de vista que cada cadena lateral tiene un conjunto definido de aminoácidos espacialmente vecinos. (Figura 2-1)

Claramente, la composición de cadenas laterales de estos entornos es diferente. La elección de los triptofanos como ejemplo no es casual. Existe una gama amplia de herramientas experimentales (algunas de ellas las veremos más adelante, en el Capítulo IV) que permite identificar en qué entorno químico se hallan estos residuos. Si es que se hallan en un entorno polar o apolar y si se encuentran en un entorno móvil, muy fluctuante o más bien rígido.

Una de las características más asombrosas es de la estructura nativa es la disminución de la movilidad de la cadena principal y de las cadenas laterales, como consecuencia del empaquetamiento ya mencionado. Si bien la estructura nativa no es rígida en un cien por cien, la dinámica general o flexibilidad se ve muy disminuida con respecto a la del estado desplegado.

Otra característica que permite identificar al estado conformacional es que, por lo general, la conformación nativa suele ser muy resistente a la acción de las proteasas. Si en cambio, exponemos una solución de proteína, parcialmente plegada o suficientemente desestabilizada a la acción de proteasas, en condiciones adecuadas, al poco tiempo de incubación observaremos degradación proteolítica, presencia de pequeños fragmentos péptidicos y la ausencia de proteína de longitud completa.

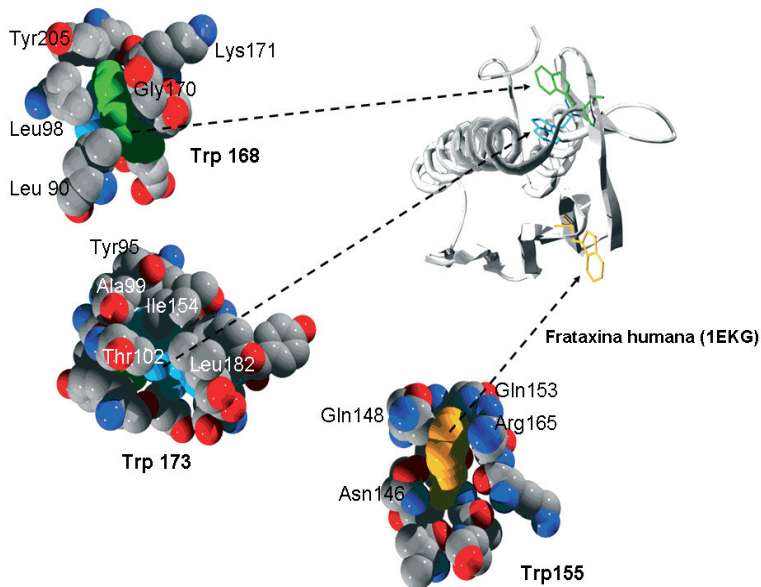


Figura 2-1. Estructura terciaria y entorno químico de los aminoácidos.

La estructura terciaria de la proteína genera asimetría en el entorno de los residuos que la constituyen. Se muestra el entorno (residuos vecinos) de los tres triptofanos (W155, W168, y W173) de la frataxina humana (1EKG).

El estado desplegado

Al estado desplegado por el contrario se lo ha caracterizado en mucha menor medida. Esto es consecuencia directa del hecho de que no se trata de una única conformación sino de un conjunto de conformaciones. Este conjunto de conformaciones adquieren estabilidad en ciertas condiciones experimentales:

- a. Concentraciones elevadas de agentes caotrópicos (desnaturalizantes) como urea o cloruro de guanidinio (desnaturalización química);
- b. Aumento de la temperatura (desnaturalización térmica);
- c. Disminución extrema de la temperatura (desnaturalización por frío).

En el caso del estado desplegado, experimentalmente, podremos medir propiedades del conjunto de conformaciones: propiedades que no corresponden a ninguna conformación particular sino al promedio de todas ellas. En general las conformaciones son altamente fluctuantes y, en principio, no deberían existir interacciones preferenciales. Sin embargo, en algunos casos se ha comprobado la existencia de estructura residual en el estado desplegado, y se admite la existencia de tendencias de la cadena principal a adquirir localmente cierto tipo de estructura secundaria.

Podemos decir sin equivocarnos que, en promedio, el estado desplegado tiene expansión máxima. Aumenta el radio hidrodinámico promedio; esto quiere decir que si intentamos modelar a los estados por medio de una esfera, el estado nativo de compactación máxima podrá ser representado con una esfera de radio, significativamente, menor que la necesaria para modelar el promedio de las conformaciones desplegadas.

50

Diferencia de Superficie accesible al solvente (ΔASA): una cuestión clave

La exposición máxima de las cadenas laterales al solvente del estado desplegado implica que existe, por lo tanto, un aumento muy significativo de superficie apolar en contacto con moléculas de H_2O .

Esta característica es sumamente importante y es una marcada diferencia con respecto al estado nativo que es, por el contrario, como se dijo, muy compacto y con mínima superficie apolar expuesta. Como veremos más adelante la diferencia entre el área de superficie accesible del estado desplegado ASA_U y la del estado nativo ASA_N es un parámetro muy importante por sus implicancias en la termodinámica del proceso de desplegado.

$$ASA_U - ASA = \Delta ASA_{NU}$$

¿Por qué es importante este cálculo?

La estabilización del estado desplegado en solución depende, en gran parte, de la solubilización de las cadenas laterales y, por consiguiente, de la superficie apolar accesible al solvente (ver efecto hidrofóbico).

La superficie expuesta al solvente del estado nativo puede modelarse a partir de las estructuras de alta resolución mediante cálculo. Los átomos expuestos serán aquellos que puedan establecer contactos con una esfera que tenga dimensiones similares a las de una molécula de H₂O (radio= 1.4 Å). Los átomos que no puedan interactuar con esta esfera que se hace rodar por la superficie del modelo proteico, estarán ocultos probablemente en el corazón de la proteína y no en su superficie.

El estado desplegado es mucho más difícil de modelar. Para comenzar, podríamos modelar su superficie expuesta asumiendo una exposición total: simplemente como la suma de sus aminoácidos constituyentes, para esto se suelen generar modelos en los que se incluye cada residuo de la proteína en un entorno de enlace peptídico mínimo.

Para esto se modelan tripéptidos, péptidos de tres aminoácidos de secuencia GXG: glicina-aminoácido modelo-glicina. Como es de esperar las glicinas de los extremos por no tener cadena lateral voluminosa impiden que la cadena lateral del residuo X (modelo) sea parcialmente desolvatada. Un modelo un poco más realista se puede obtener reemplazando las glicinas por alaninas, que cuentan ya con un carbono β (una cadena lateral mínima. En este último caso se espera una desolvatación algo mayor de cada cadena lateral.

• **Tabla 2-1.** Caracterización de las superficies accesibles para cada uno de los residuos en el contexto GXG. Se discriminan las superficies correspondientes a la cadena principal y a la cadena lateral.

Residuo	TOTAL	ASA _{Cadena principal}	Átomos de cadena lateral		
	(Å ²)	(Å ²)	ASA _{TOTAL}	ASAA _{POLAR}	ASA _{POLAR}
Ala (A)	113	46	67	67	0
Arg (R)	241	45	196	89	107
Asn (N)	158	45	113	44	69
Asp (D)	151	45	106	48	58
Cys (C)	140	36	104	35	69
Gln (Q)	189	45	144	53	91
Glu (E)	183	45	138	61	77
Gly (G)	85	85	0	0	0
His (H)	194	43	151	102	49
Ile (I)	182	42	140	140	0
Leu (L)	180	43	137	137	0
Lys (K)	211	44	167	119	48
Met (M)	204	44	160	117	43
Phe (F)	218	43	175	175	0
Pro (P)	143	38	105	105	0
Ser (S)	122	42	80	44	36
Thr (T)	146	44	102	74	28
Trp (W)	259	42	217	190	27
Tyr (Y)	229	42	187	144	43
Val (V)	160	43	117	117	0

Para poner un ejemplo concreto, si queremos estimar la superficie total de el estado desple-

gado de la frataxina humana, debemos identificar el contenido de cada aminoácido en la estructura. Para esto podemos usar la secuencia correspondiente a 1EKG, el modelo de alta resolución de la frataxina humana madura. En código de una letra la secuencia es:

1 GSHMGSLDET TYERLAEETL DSLAEFFEDL ADKPYTFEDY DVSFGSGVLT VKLGGDLGTY
61 VINKQTPNKQ IWLSSPSSGP KRYDWTGKNW VYSHDGVSLH ELLAAELTKA LKTKLDLSSL
121 AYS GKDA

A partir de la secuencia podemos establecer, por ejemplo, el contenido de cada uno de los veinte aminoácidos (ver **Tabla 2-2**).

Residuo	Contenido	
Ala (A)	8	6,3%
Arg (R)	2	1,6%
Asn (N)	3	2,4%
Asp (D)	11	8,7%
Cys (C)	0	0,0%
Gln (Q)	2	1,6%
Glu (E)	9	7,1%
Gly (G)	11	8,7%
His (H)	3	2,4%
Ile (I)	2	1,6%
Leu (L)	17	13,4%
Lys (K)	10	7,9%
Met (M)	1	0,8%
Phe (F)	4	3,1%
Pro (P)	4	3,1%
Ser (S)	14	11,0%
Thr (T)	10	7,9%
Trp (W)	3	2,4%
Tyr (Y)	7	5,5%
Val (V)	6	4,7%

Se puede usar el servidor Expasy disponible en forma gratuita en: <http://www.expasy.ch/> para realizar este cálculo en forma rápida. Herramienta Protparam <http://www.expasy.ch/tools/protparam.html>

Si multiplicamos el número de residuos de triptofano, tres (3) en nuestro ejemplo, por el ASA del triptofano en el entorno GWG habremos calculado la contribución de los residuos de triptofano a la superficie accesible total del estado desplegado ($3 \times 259 \text{ \AA}^2 = 777 \text{ \AA}^2$).

Si ahora hacemos lo mismo con cada uno de los residuos y sumamos las contribuciones habremos obtenido el área total. El ASA_U total (la superficie accesible al solvente total en nuestro modelo de estado desplegado) es $ASA_U = 20.753 \text{ \AA}^2$. Si calculamos la superficie expuesta para el estado nativo de la frataxina nativa monomérica (PDB=1EKG) obtenemos un valor de $ASA_N = 5.547 \text{ \AA}^2$. En este caso el cálculo de ΔASA_{NU} , el cambio de superficie accesible en el proceso de desplegado es de

$$\begin{aligned} \Delta ASA_{NU} &= ASA_U - ASA_N \\ &= 20.753 \text{ \AA}^2 - 5.547 \text{ \AA}^2 \\ &= 15.206 \text{ \AA}^2 \end{aligned}$$

• **Tabla 2-2.** Contenido de aminoácidos dada
• la secuencia de la cadena polipeptídica.

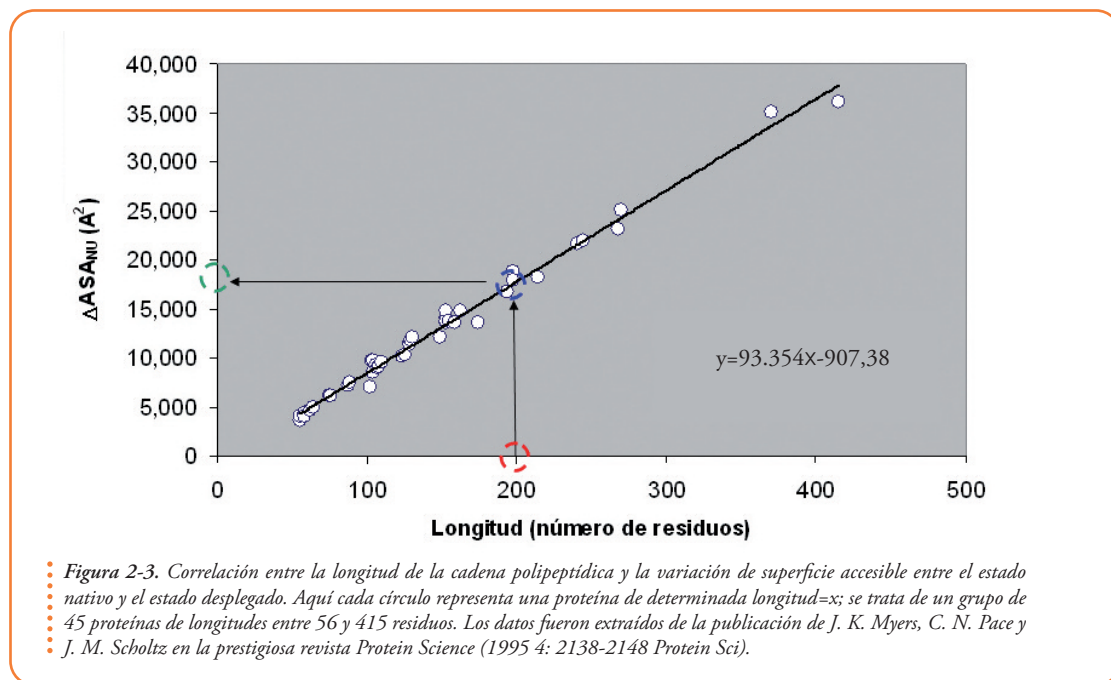
Sin embargo, este tipo de modelo sigue constituyendo un extremo donde se simula la exposición completa de cada residuo porque la secuencia de la proteína en el entorno del residuo X modula la exposición de este residuo al solvente y en este modelo el entorno son glicinas cuya cadena lateral es tan sólo un átomo de hidrógeno, muy distinto sería el panorama si los residuos vecinos fueran de cadena lateral voluminosa.

Por esto un modelo un poco más realista incluye un entorno peptídico, incluso teniendo en cuenta varios residuos a cada lado del residuo X.

En la actualidad se están realizando grandes esfuerzos por entender cómo es, estructural-

mente, el estado desplegado, qué tipo de interacciones persisten en este tipo de conformaciones. Tanto las estrategias computacionales como experimentales han hecho foco en este problema porque queda más que claro que para entender cómo se pliega una proteína debemos conocer cómo es el estado desde el que se parte en la reacción de plegado.

Como ΔSA_{NU} está directamente relacionado con la longitud de la proteína, porque en promedio las proteínas poseen una composición similar de cada uno de los veinte residuos de aminoácidos, si conocemos los valores de ΔSA_{NU} para una serie de proteínas globulares de distintas longitudes, entonces también podremos calcular, el valor de ΔSA_{NU} para la proteína de interés teniendo en cuenta únicamente su longitud (ver **Figura 2-3**).

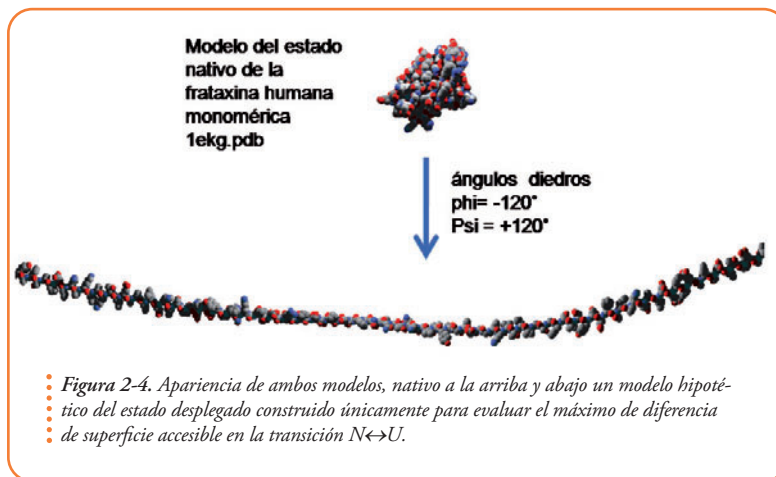


Si quisiéramos conocer el ΔSA_{NU} , por ejemplo, de una proteína de longitud $x=200$ entonces (círculo rojo en la **Figura 2-3**) deberíamos interpolar en nuestra curva de regresión (círculo azul en la **Figura 2-3**), y obtendríamos el ΔSA_{NU} (y) para la proteína de interés (círculo verde en la **Figura 2-3**). $y = 93.354 \times 200 - 907,38$

Como la longitud de la frataxina madura monomérica es de 127 residuos, en este caso el $\Delta SA_{NU} = 93.354 \times 127 - 907,38 = 10.948,6 \text{ \AA}^2$, un valor más realista y significativamente menor que 15.206 \AA^2 , el valor extremo en accesibilidad obtenido más arriba a partir de cálculos con modelos de tripéptidos GXG.

También podemos asumir, cosa que de antemano sabemos que es energéticamente improbable debido a lo inestable que resultaría esta conformación en solución, que la cadena polipeptídica puede adquirir un máximo de exposición al solvente y modelar a la cadena principal completamente extendida con ángulos torsionales $\psi = +120^\circ$, $\phi = -120^\circ$. En la **Figura 2-4** se muestra este tipo de operación realizada sobre la variante de frataxina monomérica.

Este tipo de modelos también nos permitirán obtener una estimación de la diferencia de superficie accesible al solvente entre los estados nativo y desplegado.

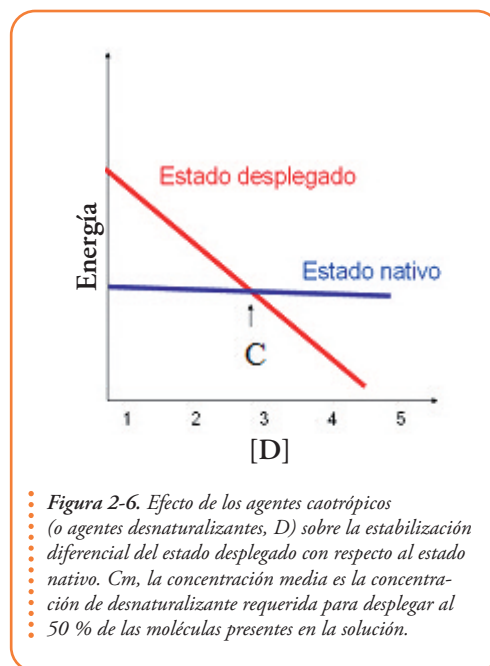


Equilibrio de desplegado $N \leftrightarrow U$.

Ya se ha mencionado que las proteínas son sistemas muy particulares porque, si bien la conformación nativa es suficientemente estable en solución acuosa, está en equilibrio con otras conformaciones que, en principio, estarán mucho menos representadas porque son, energéticamente, inestables. Estas conformaciones están, continuamente, interconvirtiéndose, sólo que las velocidades de interconversión pueden ser muy variables, e incluso despreciables en algunos casos.

En la **Figura 2-6** se muestra el efecto de un agente caotrópico (desnaturalizante, D) sobre la estabilización de los estados nativo y desplegado. Un incremento en la concentración de desnaturalizante produce una mayor estabilización (disminución de la energía) para U que para N. La pendiente de cada recta es proporcional al área expuesta al solvente en cada estado conformacional.

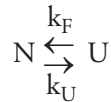
Esto se debe a que el desnaturalizante ejerce su acción como solubilizador de superficie apolar y, como hemos visto, el estado desplegado tiene una superficie accesible considerablemente mayor. Así, a bajas concentraciones de desnaturalizante U, el estado desplegado, es muy inestable con respecto a N (estado nativo) y por esto, en el equilibrio, deberíamos observar que la mayor parte de las moléculas están en conformación N, la más estable en esa condición. Por el contrario a altas concentraciones de desnaturalizante vemos que U tiene menor energía que N, por esto en la so-



lución el estado conformacional mayoritario en esta condición experimental es U. Hay un punto que es, particularmente, interesante que es la concentración de desnaturizante $[D] = C_m$ o concentración media. Esta es la concentración de desnaturizante que se requiere para que coexistan el 50% de las moléculas en conformación nativa y el 50% en conformación desplegada. En esta condición la energía de ambos estados conformacionales es igual y por esto son igualmente probables.

Las ecuaciones matemáticas que nos permiten entender la termodinámica de este proceso no son muy complejas.

Si planteamos que existe un equilibrio entre las formas N y U, en principio, este equilibrio está dado porque existe inter conversión entre las formas N y U. Las reacciones de repliegado ($U \rightarrow N$) y de desplegado ($N \rightarrow U$) pueden transcurrir con velocidades muy distintas.



Entonces podemos definir una constante de equilibrio K para cada condición experimental.

Por ejemplo en ausencia de desnaturizante, a 25°C y en condiciones determinadas de solvente (regulador de pH o *buffer*).

$$K = \frac{[U]}{[N]}$$

Podemos definir a su vez las fracciones de U y N, respectivamente, de tal forma que, como en nuestro modelo la proteína no puede estar en otra conformación que no sea N o U, si sumamos la fracción de moléculas que están completamente desplegadas (f_U) y la fracción de moléculas que están en estado nativo (f_N) entonces

$$1 = f_N + f_U$$

También debemos tener en cuenta que podemos reescribir la constante de equilibrio K como

$$K = \frac{[U]}{[N]} = \frac{f_U}{f_N}$$

En la concentración de desnaturizante $[D] = C_m$,

$$0,5 = f_N = f_U$$

Además podemos definir a la energía libre del proceso de desplegado como

$$\Delta G^{\circ}_{NU} = -RT \ln K$$

Donde K es la constante de equilibrio, R es la constante de los gases (una constante de proporcionalidad) T es la temperatura en Kelvin.

Un punto importante es que las formas N y U se inter convierten. En la concentración de desnaturizante $[D] = C_m$, la velocidad de conversión de desplegado (reacción $N \rightarrow U$) es igual a la velocidad de repliegado (reacción $U \rightarrow N$).

Por otro lado el ΔG°_{NU} , la energía libre de desplegado varía linealmente con la concentración de desnaturalizante ver **Figura 2-7**.

$$\Delta G^\circ_{NU} = \Delta G^\circ_{H_2O_{NU}} - m_{NU} [D]$$

En la **Figura 2-7**, también, se puede ver que, efectivamente, la extrapolación a concentración de desnaturalizante cero rinde un valor de el $\Delta G^\circ_{H_2O_{NU}}$. Por otro lado, como ya se mencionó, cuando la concentración de desnaturalizante es $[D] = C_m$, el valor de ΔG°_{UN} es cero, esto quiere decir que el proceso es igualmente probable en cualquiera de las dos direcciones (reacción $N \rightarrow U$ y reacción $U \rightarrow N$).

$$\Delta G^\circ_{NU} = -RT \ln K$$

$$-\frac{\Delta G^\circ_{NU}}{RT} = \ln K$$

$$e^{-\frac{\Delta G^\circ_{NU}}{RT}} = K$$

$$e^{-\frac{\Delta G^\circ_{H_2O_{NU}} - m_{NU} [D]}{RT}} = K$$

Por otro lado si pudiéramos observar alguna característica del estado nativo, S, por ejemplo, que varíe a medida que vaya aumentando la fracción de moléculas que se encuentra en estado desplegado, y si de acuerdo con los instrumentos que disponemos pudiéramos medir esta variación con precisión y exactitud, entonces podríamos seguir el proceso de desnaturalización en el equilibrio como una variación en las fracciones f_U y f_N porque la señal que se observe S_{total} será consecuencia del aporte de las moléculas que están en estado nativo, S_N es la señal que posee el estado nativo, y también de un aporte a la señal que harán las moléculas que están desplegadas; S_U es la señal del estado desplegado.

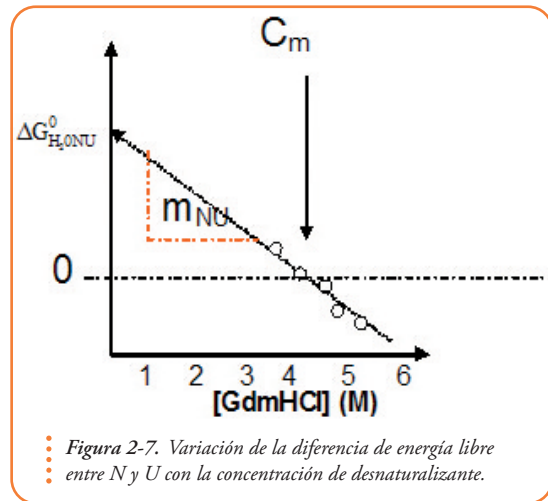
$$S_{total} = f_N S_N + f_U S_U$$

Nuestro próximo objetivo será encontrar una ecuación sencilla que nos permita dejar planteado cómo varía la señal total en nuestro experimento con respecto a la variación en la constante de equilibrio K (producto de la adición de agente desnaturalizante).

Como $f_U + f_N = 1$ podemos dividir la ecuación anterior a ambos lados por $f_U + f_N$ y nos queda

Recordemos que la energía libre (ΔG) es un parámetro termodinámico muy importante porque indica si un proceso ocurrirá espontáneamente o no: si el valor es negativo ($\Delta G < 0$), el proceso es espontáneo y no lo será si es positivo ($\Delta G > 0$); sin embargo no nos da cuenta sobre la velocidad con la que ocurriría dicho proceso.

$$\Delta G^\circ_{NU} = -RT \ln \frac{[N]}{[U]}$$



Si por ejemplo, $f_U = 0$, entonces $f_N = 1$ (porque la suma $f_U + f_N = 1$) y en este caso $S_{total} = 1 \times S_N$. Por el contrario si en la solución todas las moléculas están en estado desplegado ($f_U = 1$ y $f_N = 0$) la $S_{total} = 1 \times S_U$. En la concentración media, la señal será $S_{total} = 0,5 S_N + 0,5 S_U$.

$$\frac{S_{total}}{1} = \frac{f_N S_N}{f_U + f_N} + \frac{f_U S_U}{f_U + f_N}$$

Ahora, podemos pasar al denominador las fracciones (f_N y f_U) que están en el numerador

$$\frac{S_{total}}{1} = \frac{S_N}{\frac{(f_U + f_N)}{f_N}} + \frac{S_U}{\frac{(f_U + f_N)}{f_U}}$$

y usando la ecuación

$$K = \frac{[U]}{[N]} = \frac{f_U}{f_N}$$

Reemplazamos y obtenemos

$$\frac{S_{total}}{1} = \frac{S_N}{\left(\frac{f_U}{f_N} + 1\right)} + \frac{S_U}{\left(\frac{f_N}{f_U} + 1\right)}$$

y obtenemos

$$\frac{S_{total}}{1} = \frac{S_N}{(K+1)} + \frac{S_U}{\left(\frac{1}{K} + 1\right)}$$

Si ahora multiplicamos el segundo término en numerador y denominador por K (cosa que como es obvio no altera la igualdad)

$$\frac{S_{total}}{1} = \frac{S_N}{(K+1)} + \frac{S_U}{\left(\frac{1}{K} + 1\right)} \times \frac{K}{K}$$

entonces,

$$\frac{S_{total}}{1} = \frac{S_N}{(K+1)} + \frac{S_U K}{(1+K)}$$

y finalmente

$$S_{total} = \frac{S_N + S_U K}{(1+K)}$$

Si en esta ecuación reemplazamos K , la constante de equilibrio por

$$e^{-\frac{\Delta G^{\circ}_{H_2O_{NU}} - m[D]}{RT}} = K$$

Entonces conociendo S_N y S_U , ambos son datos que surgen de los experimentos que realizamos y son las señales que medimos en el laboratorio para la proteína, S_{total} dependerá de tan sólo dos parámetros m_{NU} (la pendiente de la recta ΔG°_{NU} en función de la concentración de desnaturalizante) y $\Delta G^{\circ}_{H_2O_{NU}}$ la diferencia de energía libre en ausencia de agente desnaturalizante, que es la ordenada al origen

después de extrapolar linealmente (ver **Figura 2-7**).

Así, nos ha quedado definida la señal esperada para cada concentración de desnaturalizante y, si ajustamos nuestro modelo a los datos experimentales obtenidos, podremos entonces encontrar estos dos parámetros que definen a la estabilidad de la proteína.

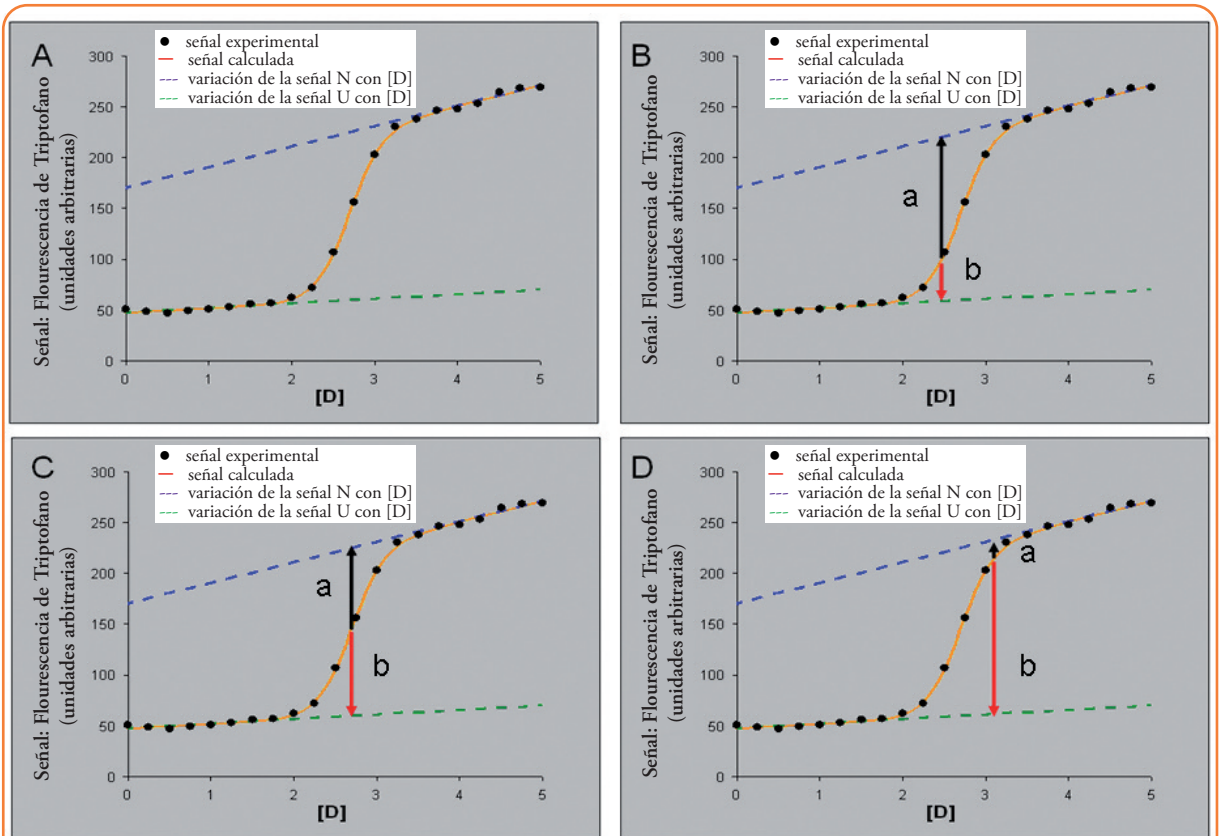
Como las señales S_N y S_U suelen cambiar con una variación en la concentración de agente caotrópico podemos reescribir a S_N y S_U como

$$S_N = S_{N_{H_2O}} + m \times [D]$$

$$S_U = S_{U_{H_2O}} + m \times [D]$$

Esto no es ni más ni menos que asumir que si es que hay un cambio en la señal de alguno de los estados conformacionales, este cambio tiene una dependencia lineal con respecto a la concentración de desnaturalizante. Estas variaciones, por lo general, no son producto de un cambio estructural importante sino son consecuencia de pequeñas alteraciones locales como producto interacciones con el desnaturalizante ([D]). En la **Figura 2-10** se muestra un ejemplo.

Se trata de la desnaturalización de la tiorredoxina de *E. coli* seguida por intensidad de fluorescencia de triptofanos. Esta es tan sólo una de las tantas señales que podemos utilizar.



••• **Figura 2-10.** Desnaturalización de tiorredoxina de *E. coli*.

En el panel (A) de la Figura 2-10, se muestran los datos experimentales (●) y en naranja el resultado de ajustar el modelo de dos estados ($N \leftrightarrow U$) a estos datos (—). En azul (---) se muestra un ajuste lineal al cambio en la señal del estado desplegado con la concentración de desnaturalizante: como se puede ver en este experimento, por encima 4.0 M de desnaturalizante las moléculas están desplegadas ($f_U \sim 1$) y sin embargo la señal sigue aumentando linealmente. En verde (---) se muestra un ajuste lineal al cambio en la señal del estado nativo con la concentración de desnaturalizante: en este caso por debajo de 1.5 M de desnaturalizante a pesar de que la mayor parte moléculas están en estado nativo ($f_N \sim 1$), la señal aumenta linealmente, este efecto no se debe a un cambio conformacional global sino a pequeñas perturbaciones locales.

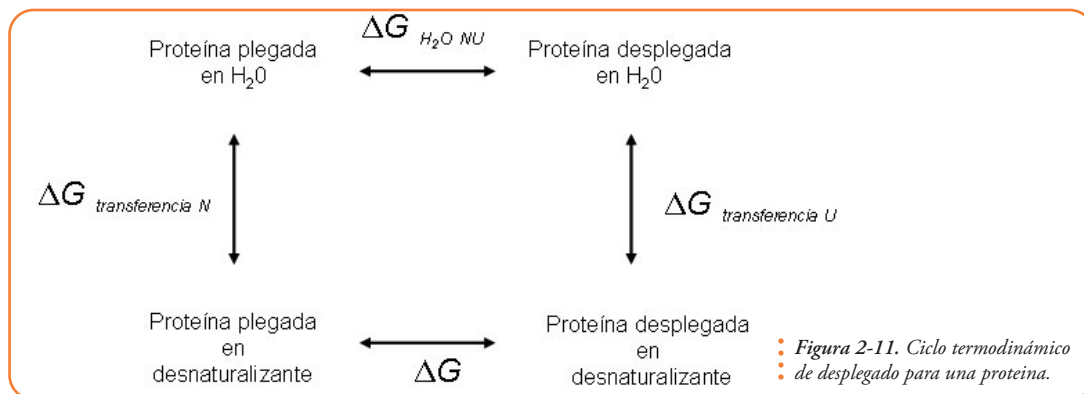
En los paneles B, C y D se muestran tres situaciones $f_N = a/(a+b)$ y $f_U = b/(a+b)$. En el panel B, la fracción de moléculas que se encuentra en estado desplegado es baja: $a/(a+b) \gg b/(a+b)$. En el panel C, $a/(a+b) = b/(a+b)$, se trata de la concentración media ($f_N = f_U$). Por último en el panel D, $a/(a+b) \ll b/(a+b)$, en este caso la mayor parte de las moléculas se hallan desplegadas.

Si nos detenemos por un momento en este experimento de desnaturalización química, podemos observar una de las principales características de las proteínas nativas: la transición entre N y U ocurre en un rango bastante estrecho de concentraciones de desnaturalizante. En este caso en el rango de 2,0 a 3,0 M. Este cambio, relativamente, abrupto está relacionado con la cooperatividad de la transición.

El término cooperatividad indica que las estabildades de las interacciones que mantienen a una proteína estructurada dependen unas de otras; por esto, cuando un cierto número de estas interacciones desaparecen, la proteína se desensambla, se despliega.

La pendiente m_{NU} es una constante de proporcionalidad (ver ecuación) y está relacionada con el cambio de superficie accesible al solvente de cada grupo de la proteína en la transición $N \leftrightarrow U$.

Para analizar esta relación podemos imaginar un ciclo termodinámico (Figura 2-11)



En este ciclo las distintas conformaciones tendrán estabilidades características en cada condición de solvente (agua o desnaturizante). En definitiva estas estabilidades determinarán la población de cada conformación presente en la solución. Podemos escribir

$$\Delta G^{\circ}_{NU} - \Delta G^{\circ}_{H_2O_{NU}} = \Delta G^{\circ}_{\text{transferenciaU}} - \Delta G^{\circ}_{\text{transferenciaN}}$$

Aquí podemos imaginar que $\Delta G^{\circ}_{\text{transferenciaU}}$ y $\Delta G^{\circ}_{\text{transferenciaN}}$ son los valores de energía libre para la transferencia de la macromolécula en la conformación desplegada y en la conformación plegada, respectivamente.

La diferencia $\Delta G^{\circ}_{\text{transferenciaU}} - \Delta G^{\circ}_{\text{transferenciaN}}$ depende, únicamente, de la exposición de grupos químicos en la transición de N a U. Así podemos reescribir la ecuación de arriba como

$$\Delta G^{\circ}_{\text{transferenciaU}} - \Delta G^{\circ}_{\text{transferenciaN}} = \sum_i \alpha_i \times n_i \times \delta G_{\text{transferencia},i}$$

Donde n es el número total de grupos de un tipo (i) en la proteína, por ejemplo, el número de cadenas laterales de leucinas en la secuencia proteica) $\delta G_{\text{transferencia},i}$ es el valor de energía libre de transferencia del grupo de tipo i, en este caso el valor de energía libre para la transferencia del residuo de leucina desde H_2O a desnaturizante. Por último el parámetro α_i representa el cambio de exposición al solvente promedio para todos los grupos de tipo i expresado como un valor entre 0 y 1; en nuestro ejemplo, como las leucinas son residuos apolares y suelen ser residuos del corazón de la proteína el cambio de accesibilidad promedio será cercano a $\alpha_i = 1$. Por el contrario los residuos polares o cargados como el ácido aspártico (Asp) son residuos que están generalmente ubicados en la superficie de la macromolécula y, por esto, el cambio de accesibilidad suele ser bajo $\alpha_i < 1$ e incluso $\alpha_i = 0$.

Nuevamente podemos hacer algunas simplificaciones y la ecuación como la forma

$$\Delta G^{\circ}_{\text{transferenciaU}} - \Delta G^{\circ}_{\text{transferenciaN}} = \Delta G^{\circ}_{NU} - \Delta G^{\circ}_{H_2O_{NU}} = \Delta \alpha \sum_i n_i \times \delta G_{\text{transferencia},i}$$

Y finalmente

$$\Delta G^{\circ}_{NU} = \Delta G^{\circ}_{H_2O_{NU}} + \Delta \alpha \sum_i n_i \times \delta G_{\text{transferencia},i}$$

La simplificación que hemos hecho es suponer que todos los grupos sufren el mismo cambio en la accesibilidad promedio $\Delta \alpha$. Si ahora tenemos en cuenta la ecuación

$$\Delta G^{\circ}_{NU} = \Delta G^{\circ}_{H_2O_{NU}} - m_{NU} [D]$$

Entonces

$$- m_{NU} [D] = \Delta \alpha \sum_i n_i \times \delta G_{\text{transferencia},i}$$

La disminución en m_{NU} está relacionada, linealmente, con una disminución en el cambio de accesibilidad promedio de los grupos de la proteína.

Como se mencionó el ΔASA_{NU} puede calcularse a partir de modelos de estado nativo y desplegado. Con este dato y los valores tabulados de energía libre para la transferencia de grupos a soluciones que contienen agente desnaturizante ($\delta G_{\text{transferencia},i}$) podemos calcular el valor del m_{NU} teórico esperado para la desnaturización completa de la proteína que estamos estudiando.

Un valor experimental bajo de este parámetro ($m_{NU \text{ experimental}} < m_{NU \text{ teórico}}$) podría implicar por lo menos dos cosas:

a. el proceso de desnaturalización que observamos no es completo, ya que el cambio de superficie accesible que tiene aparejado es menor que el esperado, dicho en otras palabras, nuestro estado final no es el estado desplegado;

b. como segunda posibilidad el estado nativo, la estructura o modelo que utilizamos para nuestro cálculo, podría no ser aplicable si en realidad nuestra proteína se encontrara en ausencia de desnaturalizantes en un estado parcialmente plegado (no nativo) con mayor superficie expuesta, luego si bien la transición se produce hacia el estado desplegado (completamente desplegado) involucra un menor cambio de superficie accesible que el esperado porque parte de un estado con mayor superficie expuesta.

Como se ve, ambas posibilidades incluyen la existencia de un estado conformacional que no es ni el estado desplegado ni el nativo. Este tema lo retomaremos casi inmediatamente, pero es interesante ver antes, que el valor de m_{NU} teórico puede calcularse en forma aproximada muy rápidamente.

Como se mostró, la longitud de la proteína se correlaciona muy bien con el ΔASA_{NU} , es por esto que se espera que también estén bien correlacionados los valores de longitud de cadena polipeptídica con los valores de m_{NU} (Figura 2-12).

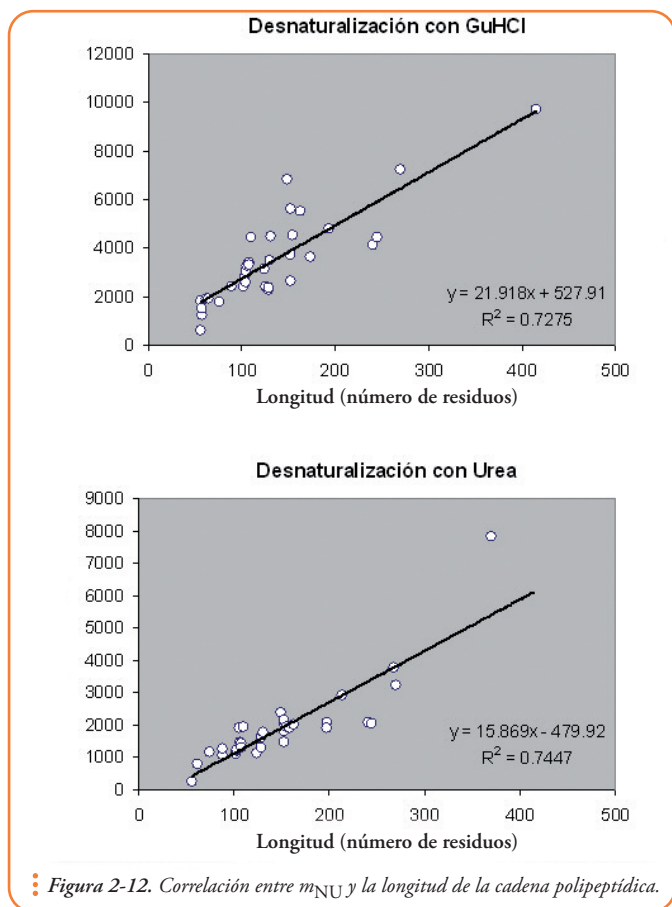


Figura 2-12. Correlación entre m_{NU} y la longitud de la cadena polipeptídica.

Utilizando el mismo juego de proteínas globulares que usamos anteriormente para correlacionar ΔASA_{NU} con longitud podemos ver que los valores de m_{NU} obtenidos experimentalmente (por numerosos investigadores) de desnaturalización con agentes caotrópicos (GuHCl o urea) también correlacionan bien con la longitud de la cadena.

Esta correlación nos permite obtener, muy rápidamente, el valor esperado de m_{NU} para procesos de desnaturalización completos utilizando, únicamente, como dato la longitud de la cadena polipeptídica. Así, el valor teórico de m_{NU} para la variante acortada monomérica de fra-

taxina humana (120 residuos) es de $1,4 \text{ kcal/mol} \times M$ para la desnaturalización en urea.

¿Cómo se lleva a delante un experimento de desplegado (y plegado) en el equilibrio?

No es difícil. Para comenzar necesitamos, en primer lugar, una buena cantidad de la proteína que queremos estudiar, típicamente unos miligramos. Esta debe estar relativamente pura (pureza >95%) porque sino lo está, estaremos examinando propiedades que nada tienen que ver con la macromolécula de interés sino con los contaminantes. Debemos tener algún tipo de equipamiento que nos permita medir alguna señal asociada al cambio conformacional ($N \leftrightarrow U$). Por ejemplo una medida del contenido de estructura secundaria o terciaria en la muestra.

*El valor experimental también es $1,4 \text{ kcal/mol} \times M$ para la desnaturalización en urea. Para el caso de la tiorredoxina de *E. coli*, el valor de m_{NU} experimental es $m_{NU}=2,95 \text{ kcal/mol} \times M$, para la desnaturalización en GuHCl , mientras que el valor teórico obtenido por correlación para una proteína de longitud= 108 residuos, es de $m_{NU}^{\text{teórico}}=2,90$. Podemos entonces concluir en ambos casos que la transición observada es $N \leftrightarrow U$ y podemos inferir del resultado experimental que existe una desnaturalización completa de la cadena polipeptídica produciéndose entonces un cambio completo superficie accesible al solvente.*

Debe quedar claro que, en este experimento, estaremos observando indirectamente al conjunto de las moléculas y no a una molécula aislada y que lo haremos a través de la observación de alguna o algunas de sus propiedades.

Además necesitaremos reactivos químicos de alta pureza para preparar el *buffer* regulador del pH (un *buffer* muy común es por ejemplo el fosfato de sodio) y nuestro equipo deberá estar conectado a un baño termostático que sea capaz de mantener la temperatura en un valor constante y deseado durante todas las mediciones. También será fundamental controlar estos parámetros durante el tiempo que le lleve a las muestras llegar al equilibrio. Temperatura y pH son dos variables muy importantes en nuestro experimento que deberemos controlar con mucha precisión. Más adelante veremos qué variaciones en la temperatura pueden producir cambios en la estabilidad de los estados conformacionales.

Antes de hacer el experimento *(por ejemplo, el que se muestra en la Figura 2-10 de desnaturalización de la tiorredoxina de *E. coli*)* debemos preparar una planilla en la que apuntaremos los volúmenes de cada uno de los componentes de nuestra solución:

- H_2O
- Buffer
- Proteína
- GuHCl

Para esto debemos, por ejemplo, tener en cuenta el volumen mínimo que necesitamos para realizar cada medición (en nuestro ejemplo el volumen mínimo fue $2.000 \mu\text{L}$). Este volumen estará dado por el tipo de equipo que disponemos y tendrá consecuencias

importantes: definirá cuánta proteína gastamos en el experimento. Debemos preparar una solución madre de proteína que podremos disolver en el *buffer* y que utilizaremos en nuestro experimento. Alternativamente, si es estable en agua la podríamos disolver en agua. De esta solución madre (cuya concentración es 355 μM) pondremos 40,1 μL por tubo (así la concentración final de proteína (7,1 μM) en cada tubo será, en principio, idéntica. Prepararemos una solución madre de *buffer* y agregaremos en todos los tubos un volumen tal que la concentración final sea constante. El *buffer* además de regular el pH de la solución aportará una serie de iones a la solución (cargas positivas y negativas) que podrían, en principio, ser importantes, En nuestro caso es *buffer* fosfato de sodio, pH 7,0 y la concentración final es 25 mM. Agregaremos el volumen de H_2O necesario para completar 2.000 μL finales en cada tubo. También prepararemos una solución muy concentrada de desnaturalizante $[\text{GuHCl}] = 8,0\text{M}$ en este caso) y agregaremos volúmenes de tal forma que las concentraciones finales de $[\text{D}]$ ($[\text{GuHCl}]_F$) sean las indicadas en la columna en amarillo: 0, 0.25, 0.5..., etc (Tabla 2-4).

[GuHCl] (m)	GuHCl (μL)	Buffer (μL)	Protína (μL)	H_2O (μL)	GuHCl (μL)
0,0	0,0	250,0	40,1	1.709,9	2.000,0
0,25	62,5	250,0	40,1	1.647,4	2.000,0
0,50	125,0	250,0	40,1	1.584,9	2.000,0
0,75	187,5	250,0	40,1	1.522,4	2.000,0
1,00	250,0	250,0	40,1	1.459,9	2.000,0
1,25	312,5	250,0	40,1	1.397,4	2.000,0
1,50	375,0	250,0	40,1	1.334,9	2.000,0
1,75	437,5	250,0	40,1	1.272,4	2.000,0
2,00	500,0	250,0	40,1	1.209,9	2.000,0
2,25	562,5	250,0	40,1	1.147,4	2.000,0
2,50	625,0	250,0	40,1	1.084,9	2.000,0
2,75	687,5	250,0	40,1	1.022,4	2.000,0
3,00	750,0	250,0	40,1	959,9	2.000,0
3,25	812,5	250,0	40,1	897,4	2.000,0
3,50	875,0	250,0	40,1	834,9	2.000,0
3,75	937,5	250,0	40,1	772,4	2.000,0
4,00	1.000,0	250,0	40,1	709,9	2.000,0
4,25	1.062,5	250,0	40,1	647,4	2.000,0
4,50	1.125,0	250,0	40,1	584,9	2.000,0
4,75	1.187,5	250,0	40,1	522,4	2.000,0
5,00	1.250,0	250,0	40,1	459,9	2.000,0

• Tabla 2-4. Diseño experimental para un experimento de desplegado en el equilibrio.

Por ejemplo, si tenemos que preparar 2.000 μL de una solución de GuHCl 3,0 M a partir de una solución de GuHCl 8,0 M usando la cuenta $V_f C_f = V_i C_i$ nuestro resultado es $(2.000 \mu\text{L} \times 3,0 \text{ M}) / 8 \text{ M} = V_i = 750 \mu\text{L}$.

Una vez incubadas las muestras suficiente tiempo como para que lleguen al equilibrio pode-

mos realizar nuestras mediciones. Este tiempo de incubación es crucial porque si la incubación es corta las muestras no habrán llegado a una situación de equilibrio. Este tiempo depende de las velocidades de plegado y desplegado de cada proteína, por esto una buena medida es, si no conocemos estas velocidades, incubar unas cuantas horas las muestras, en nuestro caso las muestras se incubaron por 3 horas a temperatura del laboratorio 20-25 °C. Más adelante, en el Capítulo IV veremos cómo se realizan las *medidas de contenido de estructura*.

¿Estado nativo o estado desplegado?

¿Cuáles son los motivos generales que determinarán la estabilización del estado nativo o la del estado desplegado?

Podemos hacer un resumen, seguramente no del todo completo. Existen contribuciones que tienen que ver con la entropía, con el orden y el desorden y otras que tienen que ver con la entalpía, la energía asociada con las interacciones intra e intermoleculares. Es importante tener en cuenta que las contribuciones entrópicas y entálpicas operan, simultáneamente, y unas se cancelan con otras de acuerdo con la ecuación

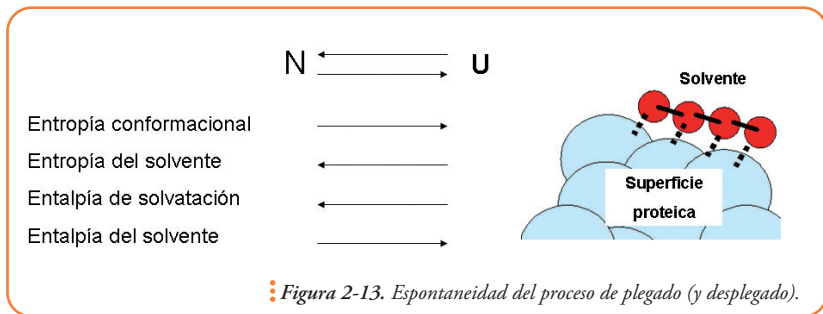
$$\Delta G^{\circ}_{NU} = \Delta H_{NU} - T\Delta S_{NU}$$

Las proteínas no son sistemas extremadamente estables por lo que el resultado de esta cuenta es, por lo general, un valor pequeño de entre 2 y 15 kcal mol⁻¹. Es muy importante tener en cuenta este dato porque, por ejemplo, la energía asociada a la formación de un puente de hidrógeno puede ser de unas 2 kcal mol⁻¹. Esto quiere decir que si eliminamos un par de puentes de hidrógeno por medio de una mutación puntual, por ejemplo a través del reemplazo de una serina por una alanina o incluso por una cisteína, podríamos desestabilizar por completo la estructura nativa. Volveremos a este punto más adelante cuando nos ocupemos de rediseñar una proteína.

Sabemos que los procesos que produzcan un aumento de entropía considerable serán favorecidos. Así, la diferencia enorme entre el número de conformaciones permitidas en el estado desplegado (U) con respecto al estado nativo (N) favorecería la estabilización de U. Sin embargo el ordenamiento de las moléculas de agua sobre la superficie accesible es mayor en el estado desplegado, porque expone más superficie, favoreciendo la estabilización del estado nativo. Por esto en el proceso de plegado deben expulsarse una gran cantidad de moléculas de agua ubicadas en la superficie proteica de U con un concomitante aumento en la entropía del solvente.

Por otro lado existe una contribución que tiene que ver con la energía de las interacciones, con la entalpía. Las moléculas de agua que solvatan la superficie de la proteína interaccionan desfavorablemente con la superficie hidrofóbica de la macromolécula por lo que se favorece la desestabilización del estado desplegado o la estabilización del estado nativo (entalpía de solvatación). Sin embargo, esas moléculas de agua interaccionan muy favorablemente entre sí generando una capa con una disposición que maximiza los puentes de hidrógeno entre moléculas de agua vecinas estabilizando el estado desplegado (entalpía del solvente).

El resultado de la sumatoria de estas contribuciones en términos de la energía libre (ΔG_{NU}), como ya se ha mencionado, permitirá definir la espontaneidad del proceso de plegado. Esto se muestra esquemáticamente en la **Figura 2-13**.



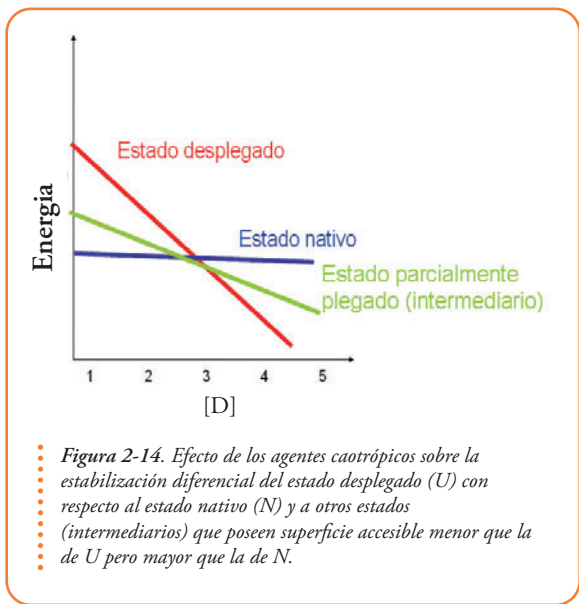
Las esferas rojas representan moléculas de solvente interactuando favorablemente entre sí (—) y desfavorablemente con la superficie proteica apolar (----).

¿Otros estado? Estados intermedios y parcialmente plegados

Las proteínas se pliegan lo suficientemente rápido como para hacer difícil la acumulación de estructuras parcialmente plegadas. Sin embargo, en algunos casos, existen barreras energéticas lo suficientemente altas que permiten la acumulación transitoria de intermediarios con estructuras dinámicas y fluctuantes. A veces estos intermediarios son estructuras incorrectamente plegadas, y para que la reacción de plegado prosiga deben volver a desplegarse. Pero, también, se observan intermediarios con estructura nativa parcial, sugiriendo un mecanismo de plegado jerárquico (ver más adelante, modelo jerárquico del proceso de plegado).

Bajo ciertas condiciones experimentales, en presencia de agentes caotrópicos, cloruro de guanidinio, urea o tiocianato de guanidinio, se observan intermediarios o estados, parcialmente, plegados en equilibrio. Las razones termodinámicas que permiten la observación de estados intermedios en el equilibrio se resumen en el modelo mostrado en la **Figura 2-14**. Aquí se ve que, para ciertas concentraciones de desnaturalizante, la estabilidad de los estados parcialmente plegados (intermediarios) es mayor que la del estado nativo y que la del estado desnaturalizado ya que la energía es menor en ese rango de concentraciones en el que el trazo verde está por debajo del azul (nativo) y rojo (desplegado).

Es importante destacar que los agentes caotrópicos aumentan la solubilidad de las cadenas laterales por lo que confieren mayor estabilidad al estado desplegado. Se establece de esta manera una dependencia funcional, cuantificable con relativa facilidad, entre la estabilidad relativa de los distintos estados conformacionales y la concentración del agente caotrópico. Como es de esperar, la dependencia está determinada por el área expuesta al solvente que



presenta en promedio cada estado. La pendiente de cada recta es proporcional al área expuesta al solvente en cada estado conformacional. Un estado parcialmente plegado podría tener una superficie accesible al solvente intermedia entre el estado desplegado y el nativo.

En el equilibrio, el factor limitante para la caracterización estructural es la heterogeneidad de conformaciones y la velocidad de inter conversión entre las mismas sin embargo hoy en día existen técnicas basadas en la resonancia magnética nuclear (ver más adelante) que permiten estudiar este tipo de conformaciones con gran detalle.

Glóbulos fundidos

Los glóbulos fundidos o *molten globules* son estructuras parcialmente plegadas que se observan en el equilibrio bajo ciertas condiciones moderadamente desestabilizantes: solventes con pH ácido, concentraciones bajas de caotrópicos, incrementos en la temperatura, y particularmente, eliminación de interacciones por mutagénesis, o por acortamiento de la cadena polipeptídica. Se trata de conjuntos de conformaciones expandidas con respecto al estado nativo, pero compactas con respecto al estado desplegado. Son sistemas poco cooperativos que suelen contener estructura secundaria, pero no interacciones terciarias fijas ni un buen empaquetamiento de cadenas laterales. Los glóbulos fundidos están estabilizados por factores entrópicos, como el aumento de libertad conformacional de las cadenas laterales. Además, en ellos, el área hidrofóbica es mayor que en el estado nativo.

Experimentalmente, se ha determinado que el reactivo (1,8)-anilinaftalenosulfonato (ANS) interacciona con las superficies hidrofóbicas del glóbulo fundido, siendo esta interacción diagnóstica de este estado. Por otro lado, se ha mostrado que, en los glóbulos fundidos, la cadena polipeptídica posee mayor flexibilidad que en el estado nativo indicando la desaparición de numerosas interacciones intramoleculares. Otra de las características relevantes de este estado conformacional es su sensibilidad a la proteólisis como consecuencia de un aumento en la flexibilidad conformacional y de la exposición de sitios de corte proteolítico. Además estos estados, por poseer superficie accesible apolar tienen una alta tendencia a agregar, a interactuar intermolecularmente y pegotarse, formando a veces agregados solubles multiméricos en los que queda oculta la superficie apolar. En otros casos estos agregados pueden precipitar.

Es importante destacar que existe una extensa variedad de glóbulos fundidos más o menos estructurados y numerosos estudios demuestran que los molten globules poseen elementos de topología nativa. Resulta muy interesante saber cómo se mantiene la topología nativa de los molten globules en ausencia de interacciones terciarias fijas. Esto podría ayudar a comprender cómo se genera la estructura terciaria. Por este motivo durante muchos años este tipo de estado conformacional atrajo la atención de numerosos investigadores.

Detección de la presencia de intermediarios en un experimento de equilibrio

Una forma relativamente sencilla de verificar la presencia de intermediarios o estructuras, parcialmente, plegadas en la solución es estudiar la no superposición de las curvas

de desnaturalización en un experimento de desnaturalización en el equilibrio. Para esto necesitaremos hacer varios experimentos en las mismas condiciones experimentales y debemos medir, por ejemplo, el contenido de estructura secundaria y estructura terciaria en las soluciones. Ya hemos visto que ciertos intermediarios poseen estructura secundaria y pueden ser compactos pero no poseen estructura terciaria definida. Si se estabilizara un intermediario en un rango de concentraciones de desnaturalizante es probable que el contenido de estructura secundaria se pierda recién a concentraciones mayores que el contenido de estructura terciaria (Figuras 2-15 y 2-16).

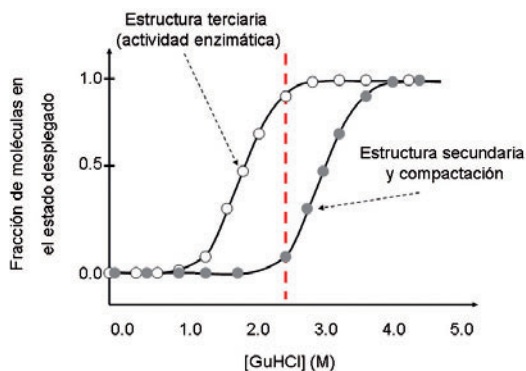


Figura 2-15. No coincidencia de las curvas de desnaturalización seguidas por medición de estructura secundaria de la proteína (círculos llenos) y por medición del contenido de estructura terciaria (círculos vacíos).

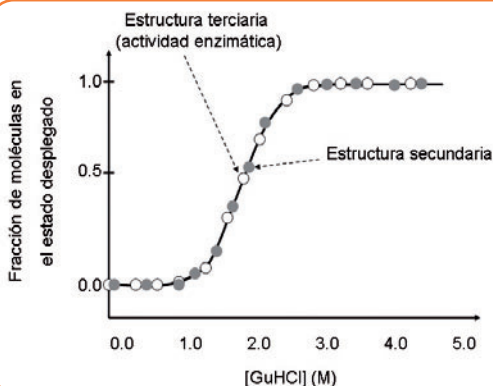


Figura 2-16. Ausencia de estados parcialmente plegados en el equilibrio.

Por ejemplo la actividad enzimática depende de la ubicación espacial de una red de residuos que participarán en la catálisis esa red queda definida por la disposición en el espacio de ellos, dicho de otra forma, por la estructura terciaria. Así, si evaluamos la actividad enzimática remanente, en cada uno de los tubos de nuestro experimento, estaremos siguiendo el contenido de estructura terciaria en cada muestra. Es de esperar que a altas concentraciones de desnaturalizante la proteína se desnaturalice y por lo tanto pierda su estructura terciaria y su actividad. Por otro lado existen ciertos tipos de espectroscopías (estudio de la interacción de la luz con la materia, en nuestro caso proteínas) que permiten evaluar el contenido de helices α o láminas β . Una de estas metodologías y tal vez la más comúnmente usada es la de absorción de luz circularmente polarizada (CD). Podremos evaluar entonces el contenido de estructura secundaria.

Alrededor de $[\text{GuHCl}] = 2.5\text{M}$ una línea roja en puntos muestra que en ese rango existe un tipo de estructura que no posee estructura terciaria (fracción de moléculas desplegadas $f_U \sim 1$, en función de la medición actividad enzimática) pero si estructura secundaria ($f_N \sim 0.1$, en función de la medición de estructura helicoidal por ejemplo por CD). A bajas concentraciones en cambio, ambas curvas coinciden (estado nativo) y a altas concentraciones también (estado desplegado, pérdida total de estructura). (ver figura 2-15.)

Ausencia de estados parcialmente plegados en el equilibrio. Si no se poblara ningún estado parcialmente plegado en la solución las sondas conformacionales coincidirían. Aquí todos los puntos (datos experimentales) caen sobre la misma curva. (ver Figura 2-16.)

Desplegado por temperatura

Uno de los métodos más comunes, aunque todavía no lo hemos discutido profundamente, es la desnaturalización térmica, y particularmente por aumento de temperatura, cuando cocinamos, desnaturalizamos proteínas y cuando hervimos la leche, también desnaturalizamos proteínas. La desnaturalización por frío, es un proceso muy interesante pero, por lo general, menos estudiado. Como en el caso de la desnaturalización química, por agentes caotrópicos, nos va a interesar evaluar antes que nada si el proceso de desplegado o desnaturalización por incremento de la temperatura es reversible o no. La aplicación de recursos teóricos de la termodinámica dependerá de este punto. Si la desnaturalización no es reversible nuestra descripción del proceso será mucho más acotada. La irreversibilidad en muchos casos ocurre por la existencia de otros procesos que están relacionados con el cambio conformacional por ejemplo la agregación. En este sentido procesos como la agregación eliminarán a la proteína del sistema, esta precipitará impidiendo el retorno de las moléculas a la conformación estado nativo.

Uno de los puntos más interesantes relacionado a la desnaturalización térmica es que, el estado nativo y el estado desplegado tienen capacidades caloríficas (C_P) distintas. El estado desplegado tiene una mayor capacidad calorífica. Esto quiere decir que para aumentar en 1 °C la temperatura de una solución de proteína nativa se necesita entregar una menor cantidad de energía que para aumentar en 1 °C la temperatura de una solución de proteína de la misma concentración en estado desplegado. Esto se relaciona con una de las propiedades que posee el estado desplegado de la que ya hemos hecho referencia muchas veces: el estado desplegado expone una superficie considerablemente mayor que el estado nativo. Las moléculas de agua próximas a la superficie proteica tienden a establecer interacciones entre sí, consolidando una red de puentes de hidrógeno intermoleculares. Así la energía entregada al sistema podrá ser absorbida por esta red de moléculas de agua, produciendo un aumento menor de temperatura de la solución, recordemos que el aumento de la temperatura es un indicador del aumento de la energía cinética del sistema.

Para entender este proceso, podemos pensar en el cambio de fase sólido/líquido para el H_2O que se da a una temperatura determinada a presión atmosférica a 0 °C. La energía entregada al sistema permitirá deformar y luego romper la red de interacciones existente en el hielo. Las moléculas vibran, alrededor de sus posiciones de equilibrio, cada vez con mayor amplitud. Llega un momento en el que vencen a las fuerzas de atracción que las mantienen fijas y el sólido se convierte en líquido. Las moléculas siguen interaccionando, pero ahora pueden moverse. Con este aumento de la movilidad de las moléculas de agua aumentará la temperatura del sistema.

En definitiva la reacción de desplegado $N \rightarrow U$ tiene un $\Delta C_{P UN}$ positivo porque $C_{P \text{ desplegado}} > C_{P \text{ nativo}}$. Como se ve este parámetro ($\Delta C_{P UN}$) está fuertemente relacionado con el ΔASA_{UN} ; es sobre la superficie de la proteína que se formará una capa de moléculas de agua más estructurada.

A medida que aumentemos la temperatura estas moléculas de agua se irán desplazando de la superficie e incrementarán su movilidad, esto incrementará el desorden y producirá un aumento de entropía) y favorecerá, en definitiva, la formación del estado desplegado. La entropía se define aquí como

$$\Delta S_{UN(T_2)} = \Delta S_{UN(T_1)} + \Delta S_P \ln \frac{T_2}{T_1}$$

Por otro lado la entalpía del estado desplegado aumenta, se desestabiliza el estado desplegado con la temperatura, porque las moléculas de agua ordenadas entorno a la superficie apolar, progresivamente, pierden las interacciones de puente de hidrógeno entre si a medida que el agua se desestructura.

$$\Delta H_{UN(T_2)} = \Delta H_{UN(T_1)} + \Delta C_P (T_2 - T_1)$$

Así el desplegado por temperatura, en realidad, es causado por el incremento entrópico que domina a altas temperaturas al término entálpico.

Las proteínas se despliegan cuando ΔG_{UN} , la energía libre, comienza a ser negativa, cuando $T\Delta S_{UN}$ es mayor que ΔH_{UN} , recordar que $\Delta G_{UN} = \Delta H_{UN} - T\Delta S_{UN}$

Existe una temperatura en la que el $\Delta G_{UN} = 0$. En esta temperatura $f_N = f_U$ por lo tanto la estabilidad del estado nativo es idéntica a la del estado desplegado en el experimento mostrado en la Figura 2-17 esta temperatura es $T_m = 68^\circ\text{C}$. Otro punto importante es que existe una temperatura de máxima estabilidad (en este caso en el rango de $30\text{-}45^\circ\text{C}$), a esta temperatura, la diferencia de energía entre el estado nativo y el estado desplegado es máxima.

Podemos ver que, en realidad, ΔG_{UN} en función de la temperatura es una campana que interseca dos veces en $\Delta G_{UN} = 0$, una vez como se ve en la Figura 2-17, es en el T_m y la segunda es en el T_C ; lo que suele ocurrir es que T_C en la mayoría de los casos es $T_C \ll 0^\circ\text{C}$, y por este motivo no se suele tener acceso experimental sencillo a esta parte de la curva de desplegado. A bajas temperaturas por ejemplo -20°C a -80°C las proteínas se despliegan. *En el laboratorio entonces tendremos que tener mucho cuidado al intentar conservar una solución de proteína. Muy probablemente no podremos conservarla en una heladera a -20°C .*

*En la Figura 2-17 se muestra un experimento de desplegado de la proteína β lactamasa del organismo *Bacillus licheniformis* en función de la temperatura. La transición entre los dos estados conformacionales nativo y desplegado se siguió por dicroísmo circular, una técnica que, como veremos, permite estudiar el contenido de estructura secundaria. En la Figura 2-17 se ve que hasta aproximadamente 50°C , el contenido de estructura secundaria es relativamente constante y que se produce un brusco cambio entre 60 y 70°C , en el que se pierde cerca del 100% de la estructura secundaria.*

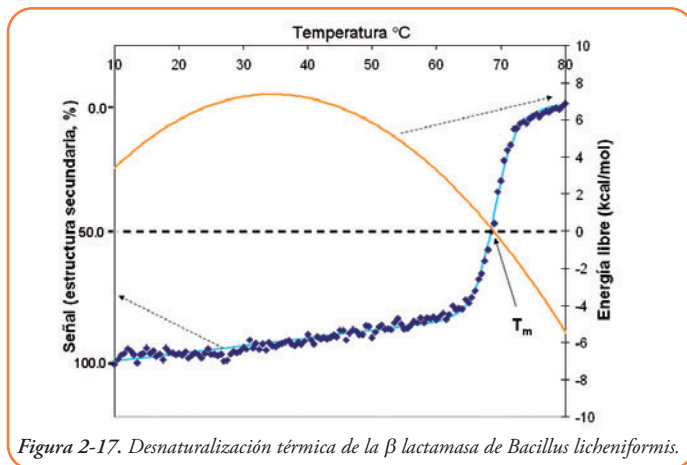


Figura 2-17. Desnaturalización térmica de la β lactamasa de *Bacillus licheniformis*.

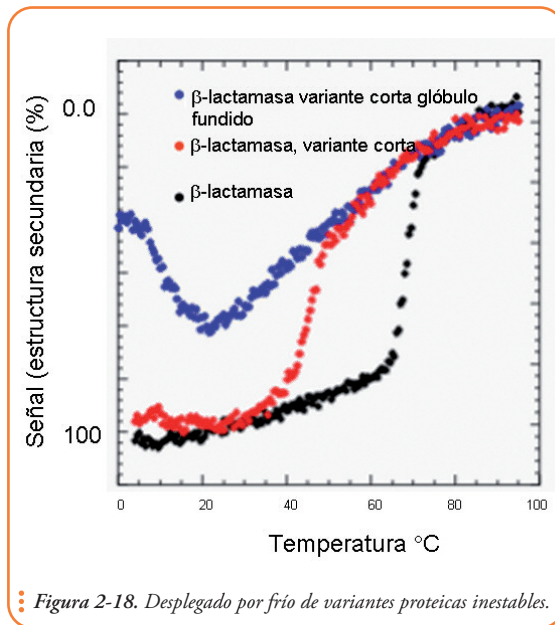
Se puede ajustar un modelo (trazo celeste) de dos estados ($U \leftrightarrow N$) a los datos experimentales (diamantes azules) de desplegado. Cómo resultado podremos extraer los parámetros termodinámicos y el ΔC_{PUN} . Luego podremos calcular la energía libre, ΔG_{UN} , en función de la temperatura (trazo naranja). (ver Figura 2-17)

A medida que la temperatura baja, como se dijo anteriormente, las moléculas de agua sobre la superficie apolar de la proteína interaccionan entre sí y se estabiliza una red de puentes de hidrógeno optimizados por la estructura del agua a esta temperatura, esto conduce a una disminución en entalpía de desplegado. La entropía (el desorden) también disminuye porque las moléculas de agua al formarse la red se ordenan.

El cambio entálpico es favorable desde el punto de vista del desplegado ($N \rightarrow U$) y es más importante que el cambio entrópico desfavorable llevando a la proteína al estado desplegado en frío (recordar que $\Delta G_{UN} = \Delta H_{UN} - T\Delta S_{UN}$).

A temperaturas mayores, las moléculas de agua, también, formarán la red sobre la superficie de la proteína pero la estructura de puentes de hidrógeno del agua no estará optimizada sobre la superficie y, consecuentemente, la reducción en entalpía no será suficiente para compensar la pérdida entrópica. Así, en esta condición la exposición de superficie apolar será desfavorable y se estabilizará el estado nativo con respecto al desplegado.

En algunos casos cuando desestabilizamos, suficientemente, a la estructura nativa podemos explorar, experimentalmente, el desplegado por frío. En la **Figura 2-18** se muestran tres curvas de desnaturalización por temperatura:



a. el desplegado de una conformación, parcialmente, plegada tipo glóbulo fundido de una mutante de la proteína β -lactamasa (una variante corta que presenta ausencia de estructura terciaria y menor contenido de estructura secundaria a 20 °C, en azul),

b. el desplegado de la conformación nativa de esta mutante (en rojo),

c. el desplegado de la variante salvaje (sin mutación) de la β lactamasa (en negro).

Puede verse cómo cuando la estructura está suficientemente desestabilizada (el caso de este glóbulo fundido) el proceso de desnaturalización por frío transcurre a temperaturas incluso $> 0^\circ\text{C}$. Además puede observarse que el proceso de desnaturalización de este tipo de estructuras altamente fluctuantes no es cooperativo y esto se traduce en que la transición ($N \leftrightarrow U$) se lleve a cabo en un rango mucho mayor de temperaturas (20 a 70°C).

El ΔC_p y el ΔSA_{UN}

La relación entre el ΔC_p y el ΔSA_{UN} ha sido estudiada por numerosos investigadores. Por ejemplo, Murphy and Freire (1992) dan una correlación entre ambos parámetros

$$\Delta C_p = 0,45 \times \Delta ASA_{np} - 0,26 \times \Delta ASA_p$$

donde **np** es no polar y **p** es polar.

Una correlación mejor es obtenida (sin tener que discriminar entre área polar y apolar) usando la ecuación desarrollada por J. K Myers y colaboradores (Myers et al. 1995)

$$\Delta C_p = 0,19 \times \Delta ASA - (864 \times n_{s-s})$$

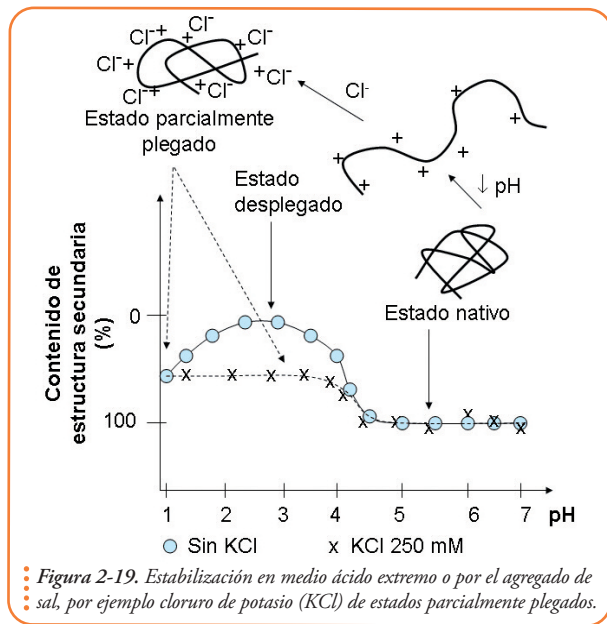
Esta ecuación contempla que, el número de puentes disulfuro existentes en las proteínas, restringirá, significativamente, la accesibilidad de la cadena principal y de las cadenas laterales.

Estas correlaciones y las correlaciones obtenidas entre la longitud en número de residuos y la diferencia de superficie accesible al solvente entre los estados nativo y desplegado se basaron en datos experimentales de diferencia de capacidad calorífica obtenidos mediante estudios calorimétricos y datos teóricos de ASA_U modelados por simulación y datos de ASA_N basados en modelos experimentales de estructuras nativas de las proteínas usadas para el cálculo.

Desplegado por pH

Otro método, relativamente, común para desplegar proteínas es a través de cambios en el pH (Figura 2-19). La acidificación de la solución de proteína (por ejemplo por el agregado de ácido clorhídrico) llevará a que se protonen todos aquellos grupos que pueden captar un protón (por ejemplo, las histidinas adquirirán carga positiva mientras que los residuos de ácido glutámico y aspártico perderán su carga negativa en este proceso). Cuando esto ocurre, los distintos segmentos de la cadena comienzan a repelerse por el accionar de fuerzas electrostática (presencia de altas densidades de cargas positivas).

Sin embargo, si al estado desplegado en medio ácido se le agregan sales capaces de apantallar, aunque sea parcialmente, las cargas positivas sobre la proteína, entonces la cadena polipeptídica podrá volver a compactarse, y algunos elementos de estructura secundaria podrán estabilizarse. Esto es lo que ocurre en el caso de los estados parcialmente plegados (glóbulos fundidos) estabilizados en medio ácido extremo o por el agregado de sal, por ejemplo cloruro de potasio (KCl).



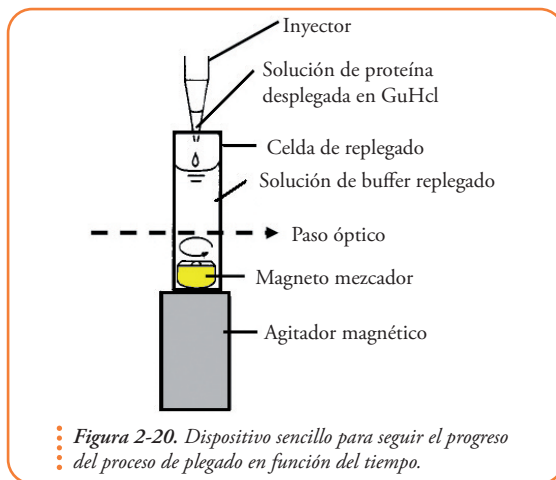
En ausencia de sales (KCl) el estado parcialmente plegado se estabiliza recién con el agregado de suficiente HCl a pH 1,0 (presencia de aniones Cl^-). En cambio si agregamos sales y luego disminuimos el pH (KCl 250 mM) el estado parcialmente plegado se estabiliza a pH mayores. (ver figura 2-19)

La reacción de plegado proteico seguida en función del tiempo.

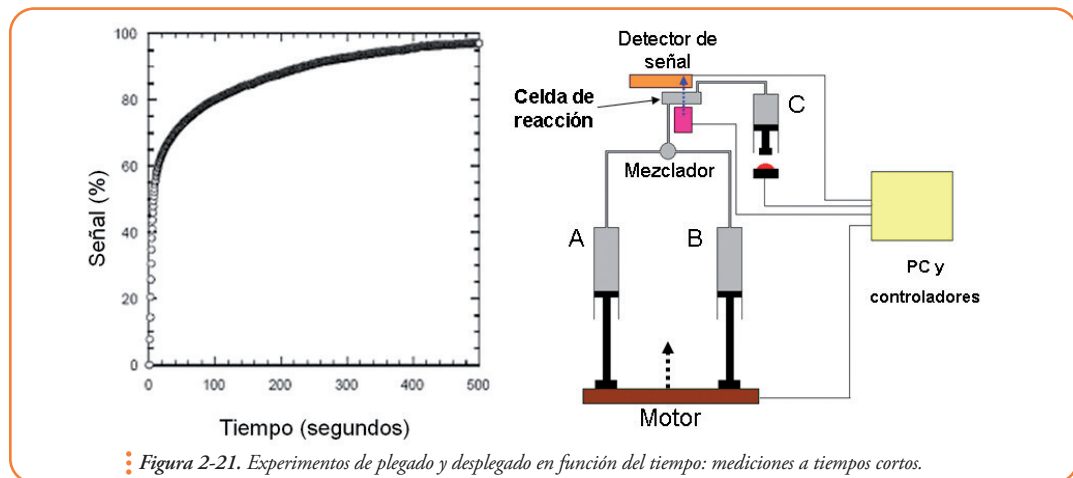
El avance de las reacciones de plegado y de desplegamiento pueden seguirse en el tiempo mediante diversas técnicas que veremos más adelante. Estas técnicas nos permitirán medir en función del tiempo la proteína replagada desplegada. En la **Figura 2-20** se muestra un ejemplo sencillo de un sistema para medir el avance de la reacción de plegado.

En este caso se prepara una solución de proteína desplegada (por ejemplo en 6.0M de GuHCl) y se inyecta un volumen de esta solución en una celda o cubeta de replagado (una cubeta de cuarzo) que tiene buffer de replagado y en el caso más simple no tiene agente desnaturizante y un magneto mezclador que permite unir, rápidamente, el volumen agregado de proteína y, por lo tanto, diluir la concentración de desnaturizante.

Demorando el menor tiempo posible, se observa cómo cambia alguna propiedad de las macromoléculas a medida que se repliegan, por ejemplo alguna propiedad óptica, es decir, cómo interactúan las macromoléculas con la luz. Por ejemplo, podemos evaluar cómo cambia en función del tiempo la absorción de luz por parte de la solución de proteína. Para esto debemos usar una celda de replagado transparente, en este caso no debe absorber luz en el rango de longitudes de onda en el que se mide. Un punto importantes es que algunas proteínas se pliegan (o despliegan) lentamente, en segundos a minutos, en cambio otras lo hacen muy rápidamente, desde milisegundos hasta micro segundos, así que dependiendo de qué proteína se trate podremos o no usar este tipo de dispositivo sencillo. En los casos en los que se pliegan muy rápidamente se debe utilizar otro tipo de equipamiento, ahora con dispositivos controlados electrónicamente que permiten inyectar, mezclar eficientemente y medir en tiempos muy cortos. Uno de estos equipos se detalla en la **Figura 2-21**.



72



Aquí en la Figura 2-21, como se ve, las cosas se complejizan un poco: tenemos tres jeringas controladas por una computadora. En la jeringa A se almacena proteína desplegada en desnaturalizante, en jeringa B solución de replegado. Cuando el motor (placa marrón) avanza los émbolos de las jeringas A y B impulsan un volumen (que se puede controlar) hacia el mezclador, aquí se mezclan ambas soluciones y la mezcla ingresa a la celda de reacción desplazando el volumen presente en ella hacia la jeringa C, el émbolo de la jeringa C (a medida que se llena la jeringa) retrocede hasta cierto punto (este volumen puede regularse) en el que se detiene por contacto (frenado de flujo), simultáneamente se dispara una señal electrónica (al oprimirse el botón rojo de frenado de flujo, que indica que la reacción ha comenzado). En este caso también se mide alguna propiedad que cambia con el replegado de la proteína a medida que la reacción avanza. En el panel de la izquierda se muestra como cambia una propiedad óptica (señal) en el proceso de replegado de la proteína β -lactamasa a medida que transcurre tiempo de reacción. Se parte del estado desplegado que tiene señal= 0%. En los primeros 100 segundos se ha adquirido más del 80% de la señal final (estado nativo). Vale la pena ver que en el proceso de replegado de esta proteína existe una etapa muy rápida, a los 20 segundos de reacción la señal ya ha tomado un valor mayor al 60% de la señal que tiene una solución en la que todas las moléculas se encuentran en estado nativo.

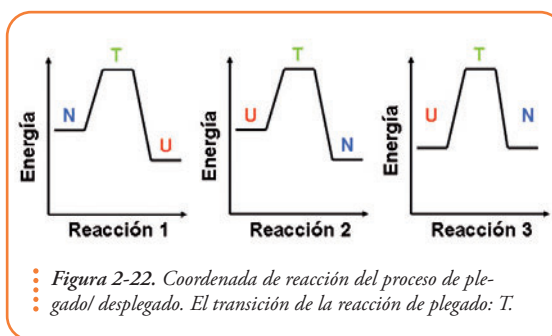
El estado de transición de la reacción de plegado

Para descubrir los principios que gobiernan el proceso de adquisición de estructura nativa es necesario, además, comprender la naturaleza de las barreras que lo bloquean y las fuerzas que lo impulsan.

Durante el proceso de plegado se generan muchas uniones no covalentes. La formación de estas interacciones limita la velocidad, imponiendo una barrera energética debido a los notables cambios en la entropía conformacional, durante el plegado se restringe la libertad conformacional de la cadena polipeptídica que adquiere una única conformación, la nativa, y de hidratación, a medida que avanza la reacción de plegado una gran cantidad de moléculas de agua que hidrataban a la cadena polipeptídica desplegada deben liberarse al seno del solvente permitiendo la compactación de la macromolécula.

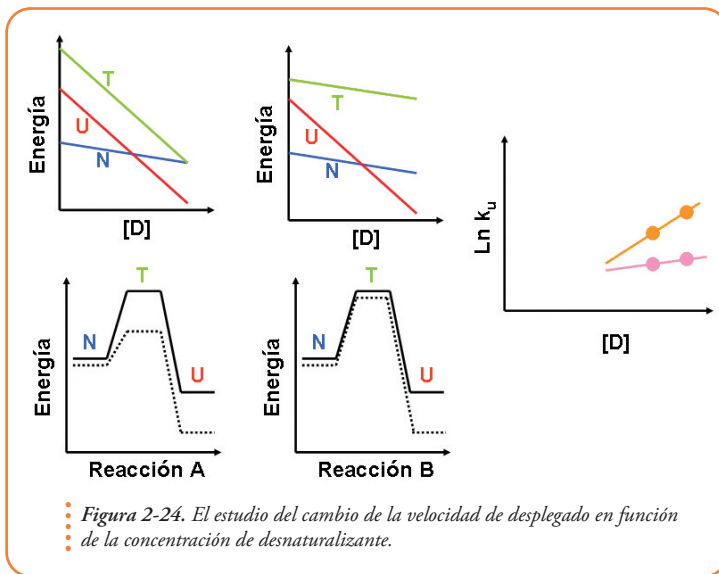
El conjunto de estructuras en formación que están en la cima de la barrera energética, y constituyen el máximo de energía en la coordenada de reacción, se conoce como estado de transición de la reacción de plegado (T^\ddagger). El estado de transición está formado por un conjunto de estructuras inestables que, por definición, tienen un 50 % de probabilidad de progresar al estado nativo.

Para el caso de la reacción 1 (Figura 2-22 panel de la izquierda) el estado desplegado es más estable que el estado nativo; en esta condición experimental, la reacción transcurrirá hacia U, lo contrario sucede en el caso de la reacción 2. Por último en la reacción 3 la estabilidad de los nativo y desplegado es idéntica (misma energía) por esto las barreras ener-



de la misma forma que para el estado desplegado porque los agentes caotrópicos, como ya mencionamos, actúan solubilizando superficie expuesta (Izquierda panel superior, en la **Figura 2-24**). Nuestro experimento se puede llevar a delante por medio de una dilución de proteína en estado nativa en un medio que contenga una concentración definida de agente desnaturante. Más adelante veremos algunos de los problemas técnicos con los que nos enfrentamos al realizar este tipo de experimentos.

Si hacemos la reacción de desnaturalización (desplegado) en dos concentraciones distintas de agente caotrópico ($[D]$), el trazo punteado representa una concentración mayor de desnaturante que el trazo continuo) las energías relativas de los estados conformacionales podrían representarse cualitativamente como en los paneles inferiores de la **Figura 2-24** (reacciones A y B). Aquí nuevamente se ve que la estabilización que sufre cada estado conformacional depende de la superficie que expone. El estado desplegado se estabiliza más que el estado nativo, y en este caso, como el estado de transición posee tanta



superficie expuesta como el estado desplegado (nuestro supuesto inicial) se estabiliza tanto como este (**Figura 2-24**, paneles superior e inferior, izquierda). Lo interesante es que en este caso como la constante de velocidad de desplegado es inversamente proporcional a la altura de la barrera NT, un aumento en la concentración de agente desnaturante incrementará la velocidad de desplegado.

Pero, sin embargo, también podemos imaginarnos que el conjunto de estructuras fluctuantes e inestables a las que llamamos estado de transición son tan compactas como el estado nativo. Ahora nuestro supuesto inicial es el otro extremo. Esto se representa en el panel de la derecha. La energía del estado de transición no variará tanto con el incremento en la concentración de desnaturante. La pendiente de la recta (verde) ahora será similar a la pendiente de la recta azul que representa el efecto del incremento de desnaturante sobre la estabilidad del estado nativo (**Figura 2-24**, panel superior centro). Como se ve el estado desplegado expone más superficie y por lo tanto se estabiliza más con una concentración de desnaturante mayor (línea de puntos). Bajo esta condición la barrera energética no cambia significativamente con un aumento de agente desnaturante. Por este motivo la velocidad de desplegado tampoco variará significativamente.

En el panel de la derecha de la **Figura 2-24** se muestra un gráfico que se conoce con el nombre de gráfico de Chevron, brazo (o rama) derecho. Aquí lo que se grafica es cómo varía la constante de velocidad de desplegado k_U con la concentración de desnaturante para una proteína que posee un estado de transición extremadamente compacto (trazo rosa) y para una proteína que posee un estado de transición extremadamente expandido (naranja). Como dijimos más arriba, si estado de transición está expandido variará significativamente la velocidad de desplegado mientras

que si es compacto el cambio en la velocidad de desplegado en función de la concentración de agente caotrópico será menor. Así analizando las pendientes de estos brazos podemos definir una de las características más importantes del estado de transición: su grado de compactación.

El análisis del brazo derecho del gráfico de Chevron es el complemento que necesitamos (Figura 2-25). La reacción que se estudia en este caso es el replegado desde el estado desplegado hasta la forma nativa en distintas concentraciones de desnaturalizante. Ahora si el estado de transición es expandido (tanto como el estado desplegado), la barrera energética que separa a U de N será siempre la misma porque U y T se estabilizan de la misma manera (Figura Panel inferior izquierda). Mientras que si el estado de transición es compacto (tan

compacto como el estado nativo), con el agregado de desnaturalizante únicamente se estabilizará U (baja su energía, Figura panel inferior centro). De esta manera variará, considerablemente, la velocidad de replegado porque al incrementarse la barrera con el desnaturalizante la velocidad disminuirá.

En el gráfico de la derecha se muestra este efecto: la dependencia de la velocidad de plegado con la concentración del desnaturalizante para dos concentraciones. En naranja para un estado de transición expandido: no hay cambio de barrera, no hay cambio en la constante de velocidad de plegado (k_F). Por el contrario, en rosa se muestra cómo un estado de transición compacto determina un incremento en la dependencia de la constante de velocidad de plegado con la concentración de agente caotrópico. El incremento en la pendiente es indicador de este proceso: mayor [D] menor k_F .

Si juntamos ambas ramas, la derecha (plegado) y la izquierda (desplegado) obtenemos el gráfico de Chevron completo (Figura 2-26).

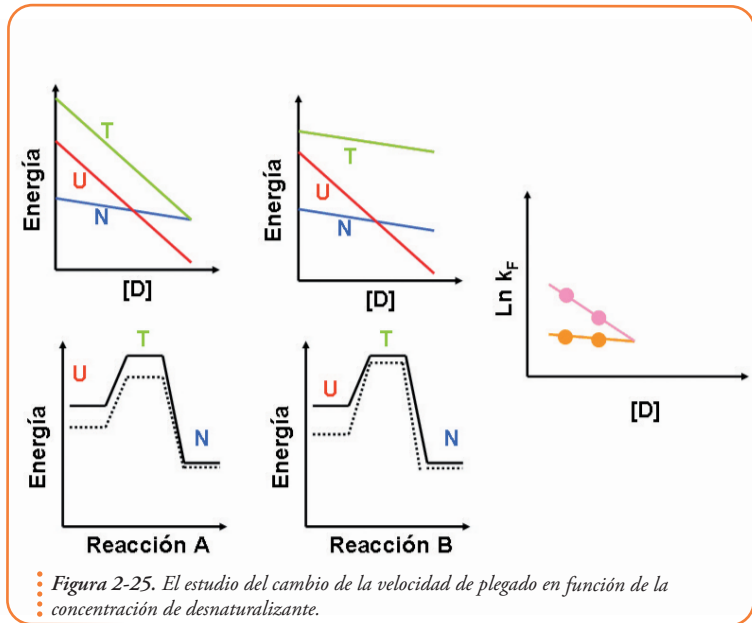


Figura 2-25. El estudio del cambio de la velocidad de plegado en función de la concentración de desnaturalizante.

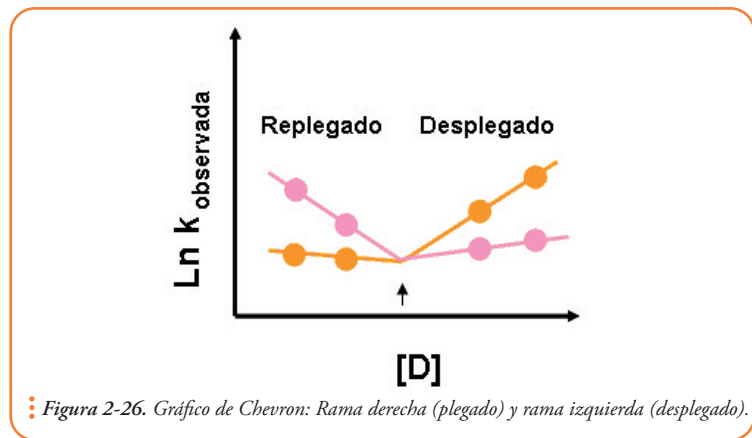


Figura 2-26. Gráfico de Chevron: Rama derecha (plegado) y rama izquierda (desplegado).

Es importante ver que, como se mencionó anteriormente, cuando las energías del estado desplegado y del estado nativo son idénticas las velocidades de plegado (k_F) y desplegado k_U serán iguales (Figura 2-22 reacción 3).

Si miramos atentamente un gráfico de Chevron, podemos ver que existe una región donde el brazo derecho toma valores semejantes al brazo izquierdo. Justamente, esto ocurre en la concentración media de agente caotrópico (C_m). En la Figura 2-27 se muestra en el panel superior un gráfico de Chevron en el que si se hacen extrapolaciones a concentración de desnaturante $[D]=0$, se obtienen las constantes de velocidad en H_2O (en ausencia de desnaturante). Recordemos que nuestros datos experimentales son los puntos naranjas. En el panel central podemos ver un experimento de despliegado en el equilibrio en el que se grafica la fracción de las moléculas de proteína que se encuentran desplegadas (a bajas concentraciones de agente desnaturante $f_U \sim 0$ (ninguna de las moléculas se ha desplegado) y a altas concentraciones $f_U \sim 1$ (cerca del 100 % de las moléculas se ha desplegado). En la concentración de desnaturante $C_m = [D]$ el 50% de las moléculas se encuentra desplegado, aquí entonces $f_U = 0.5$. La línea vertical de puntos muestra como esta $[D] = C_m$ es la región donde $k_U = k_N$ en el gráfico de Chevron (panel superior).

Una de las técnicas más poderosas de la química orgánica para probar la estructura de T^\ddagger en una reacción simple consiste en el estudio de la relación 'estructura reactividad'. Este tipo de estudio comienza con la preparación de una serie de reactivos que poseen pequeñas modificaciones químicas con respecto al reactivo original. Posteriormente, se mide el cambio en la constante de equilibrio y en las constantes de velocidad de la reacción. En base a los resultados pueden inferirse aspectos de la estructura de T^\ddagger .

Para estudiar el estado T^\ddagger del plegado

Vale la pena recordar que en el equilibrio (ver desnaturalización), la constante de equilibrio $K_{eq} = k_U / k_F$ y, a su vez, $K_{eq} = [U] / [N]$ o $K_{eq} = f_N / f_U$. Así en la concentración media (C_m) de un experimento típico de desnaturalización en el equilibrio vale $k_U = k_F$, $[U] = [N]$, $f_U = f_N$.

También vale la pena recordar que

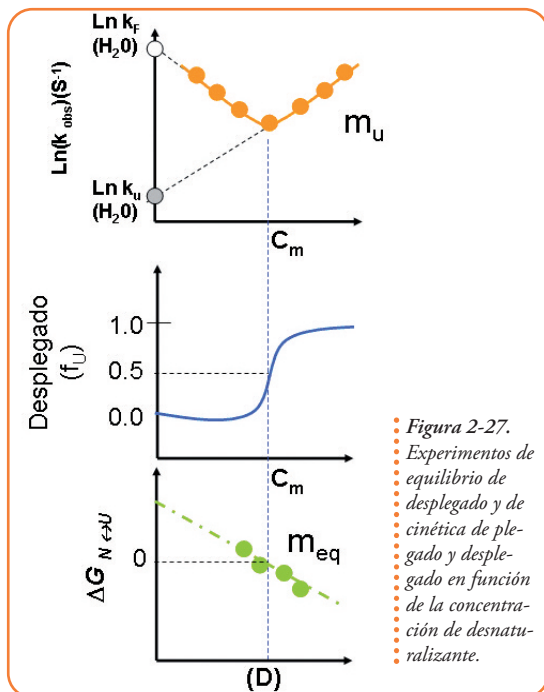
$$\Delta G^\circ_{NU} = -RT \ln K$$

$$\text{y como } K = K_{eq} = [U] / [N] = k_U / k_F$$

vale en la concentración media que

$$\Delta G^\circ = -RT \ln k_U / k_F = 0$$

(efectivamente la estabilidad de U es la misma que la de N y por esto la probabilidad de que la reacción transcurra en un sentido (plegado) es la misma que transcurra en el otro (desplegado)).



proteico, se desarrolló un procedimiento experimental basado en los mismos conceptos, pero sustentado en la ingeniería de proteínas. Este procedimiento permite medir la magnitud y la importancia de una interacción particular, y cómo esta afecta la cinética, el equilibrio y el mecanismo de plegado. La estrategia, muy sintéticamente, incluye la modificación por ingeniería genética de una interacción. Se reemplaza un residuo por otro eliminando una parte de la cadena lateral, por ejemplo leucina por alanina. También es necesario establecer el efecto de la modificación sobre los estados nativo y desplegado. Del análisis de los cambios en la magnitud de la barrera energética que separa al estado desplegado del nativo, puede inferirse la estructura del estado de transición. En efecto, cada mutación actúa como una sonda específica del grado de formación de estructura en el micro entorno del residuo modificado. Es decir que puede inferirse si la interacción medida está o no formada en T^{\ddagger} .

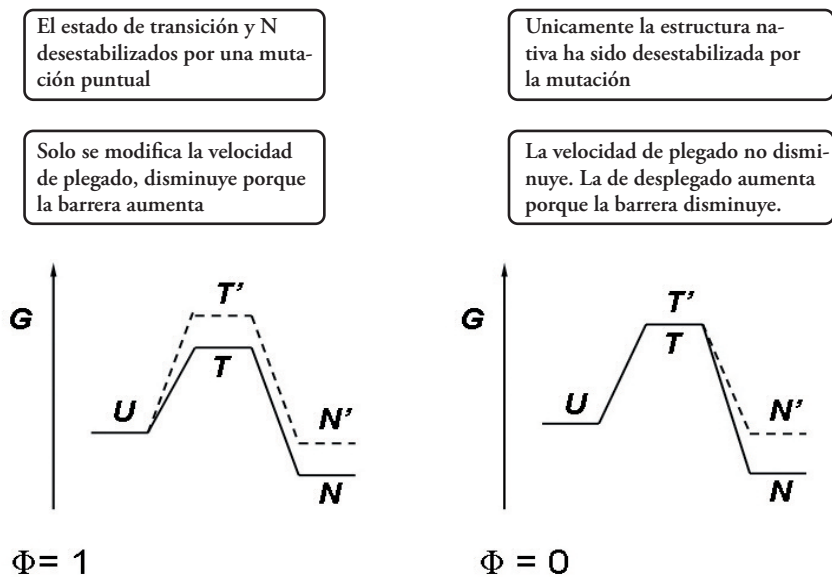


Figura 2-28. Análisis de phi de Alan Fersht para estudiar el estado T^{\ddagger} del plegado proteico. Del análisis de los cambios en la magnitud de la barrera energética que separa al estado desplegado del nativo, puede inferirse la estructura del estado de transición.

La aplicación en forma rigurosa de este tipo de procedimiento experimental está limitada a proteínas que se pliegan en dos estados ($U \leftrightarrow N$) o a 'tramos' del proceso de plegado. Por ejemplo, en la reacción de plegado $U \leftrightarrow I_1 \leftrightarrow N$ se limita al proceso $I_1 \leftrightarrow N$ que puede considerarse como una reacción de plegado de dos estados. Es importante notar que las mutaciones que desestabilicen al estado de transición tendrán un efecto sobre la velocidad de la reacción de plegado. De esta manera, más alta es la barrera energética, más lento será el proceso de plegado. La constante de velocidad de reacción k es proporcional a $e^{(-\Delta G_{TU})}$.

El valor **phi** (Φ) es un indicador de la presencia de determinada interacción en el estado de transición. Este parámetro toma el valor $\Phi=1$ cuando la interacción desestabiliza por igual a T y al estado nativo indicando que la interacción presente en el estado nativo también estaba formada en el estado de transición y contribuía con la estabilidad del mismo del mismo

modo que contribuye con la estabilidad del estado nativo. En cambio, cuando, solamente, afecta al estado nativo y no al estado de transición phi toma el valor $\Phi=0$ (Figura 2-28 y 2-29). Esto ocurre porque phi se define como

$$\phi = \frac{\Delta\Delta G_{TU}}{\Delta\Delta G_{NU}}$$

Donde $\Delta\Delta G_{NU}$ no es, ni más ni menos, que la diferencia de estabilidades de la proteína mutante con respecto a la salvaje ($\Delta G_{NU \text{ salvaje}} - \Delta G_{NU \text{ mutante}}$) y $\Delta\Delta G_{TU}$ es la diferencia de estabilidad del estado de transición entre las dos variantes, $\Delta G_{TU \text{ salvaje}} - \Delta G_{TU \text{ mutante}}$ que puede obtenerse calculando las alturas de las barreras mediante experimentos en los que se mide la velocidad de plegado.

Existen evidencias de una fuerte correlación entre la velocidad de plegado de proteínas simples (que poseen un único dominio) con parámetros que miden la complejidad topológica del estado nativo. Esta correlación sugiere que la topología podría ser determinante para alcanzar el estado de transición. Por el contrario, no hay correlaciones entre el tamaño o la estabilidad y la cinética de replegado. Este tema será abordado nuevamente más adelante. Sin embargo, resulta interesante ver que las barreras energéticas para el plegado pueden tener base entálpica

(interacciones nativas ausentes o interacciones no nativas presentes) o base entrópica (topologías complicadas). En definitiva ambos términos intervienen en el plegado de la cadena polipeptídica en el espacio. Existe un grupo de proteínas, particularmente, llamativas en las que, aparentemente, no se detecta una barrera energética a través de un estado de transición; a este tipo de plegado se lo ha llamado cuesta abajo (*down hill* en inglés). Algunos investigadores piensan que los mecanismos de plegado en dos estados ($N \leftrightarrow U$) resultarían de una falta de sincronización entre la ganancia de energía de estabilización y la pérdida de entropía conformacional (estado desplegado representado por miles de conformaciones mientras que el estado nativo estaría representado por una). Se piensa que la sincronización entre estos dos términos resulta en un plegado cuesta abajo (curva roja en la Figura 2-30).

Se caracteriza por la ausencia de una barrera energética a través de un estado de transición. Se representa con un trazo en color rojo. (ver Figura 2-30)

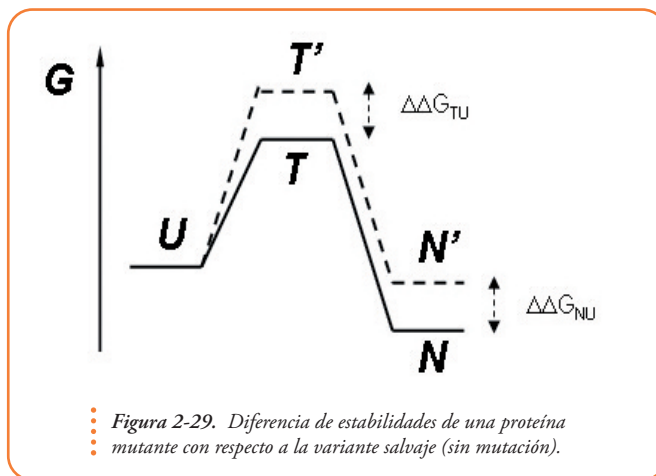


Figura 2-29. Diferencia de estabilidades de una proteína mutante con respecto a la variante salvaje (sin mutación).

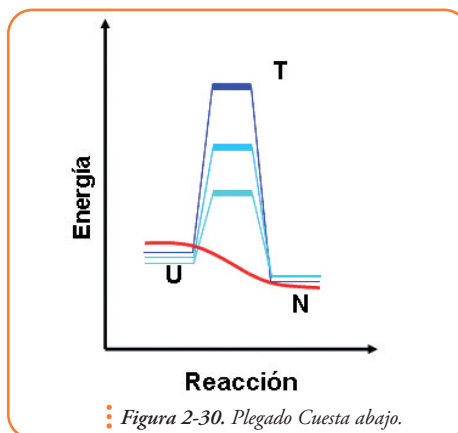


Figura 2-30. Plegado Cuesta abajo.

También es importante notar cómo varían estos diagramas en el caso de la presencia de intermediarios. En la figura se muestra un ejemplo hipotético en el que existen en la reacción de plegado por lo menos dos intermediarios (I_1 e I_2) ambos con características estructurales bien distintas. I_1 posee toda la estructura secundaria del estado nativo pero no es compacto. En cambio I_2 es casi tan compacto como el estado nativo, si bien no se han fijado aún todas las interacciones terciarias nativas. Sin embargo cuando analizamos el estado de transición de la reacción de plegado la etapa lenta limitante es la conversión de I_2 en N. La barrera más alta es efectivamente la barrera limitante de la velocidad de plegado (Figura 2-31).

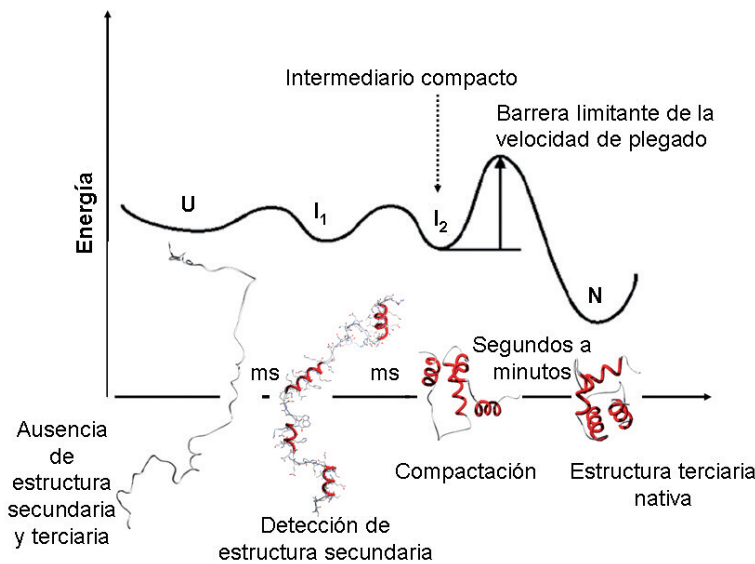


Figura 2-31. Un ejemplo hipotético que describe una ruta de plegado hacia la estructura nativa a través de varios intermediarios.

La paradoja de Levinthal

Por mucho tiempo coexistieron dos conceptos muy interrelacionados en torno al estudio del proceso del plegado de las proteínas: la paradoja de Levinthal y la supuesta necesidad de la existencia de intermediarios o vías específicas del plegado.

Levinthal dedujo, hace 30 años, que si la búsqueda de la conformación nativa desde el estado desplegado fuera un proceso aleatorio, las moléculas deberían cambiar su conformación tantas veces que sería un proceso irrealizable en tiempos razonables. Por ejemplo, si una cadena de cien residuos tiene 10100 posibles conformaciones (un promedio de diez conformaciones por residuo), y si una conformación se convierte en otra en el tiempo más corto posible (tal vez 10-13 ps), el tiempo promedio requerido para explorar todas las posibles conformaciones hasta hallar el mínimo energético sería 1.077

años. Las observaciones, por el contrario, indican que las proteínas se pliegan en tiempos muy cortos, desde unos pocos microsegundos a minutos u horas.

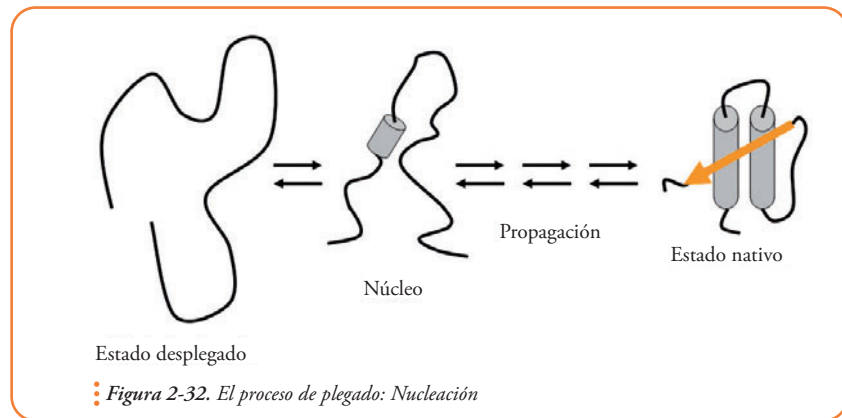
A partir de este tipo de consideraciones Levinthal propuso la existencia de vías discretas (caminos) de plegado que eviten las conformaciones no productivas (las que no conducirían hacia el estado nativo).

Modelos de plegado: Teorías para el proceso de plegado

Nucleación

En el contexto descrito anteriormente surgieron una serie de modelos sobre las rutas de plegado. El estudio de las vías de plegado debía, en teoría, revelar el mecanismo de plegado y ayudar a determinar el código por el cual la secuencia determina la estructura.

El propio Levinthal, en el año 1968, sugirió que una parte de la cadena polipeptídica podría servir como núcleo para una posterior propagación de la estructura 3D (Figura 2-32). Este modelo requiere que el núcleo sea suficientemente pequeño para que una búsqueda al azar permitiera encontrar, localmente, la estructura nativa. Una vez que el núcleo alcanza la estructura nativa se posibilita el plegado secuencial y esencialmente independiente de cada residuo.



Difusión, colisión y coalescencia

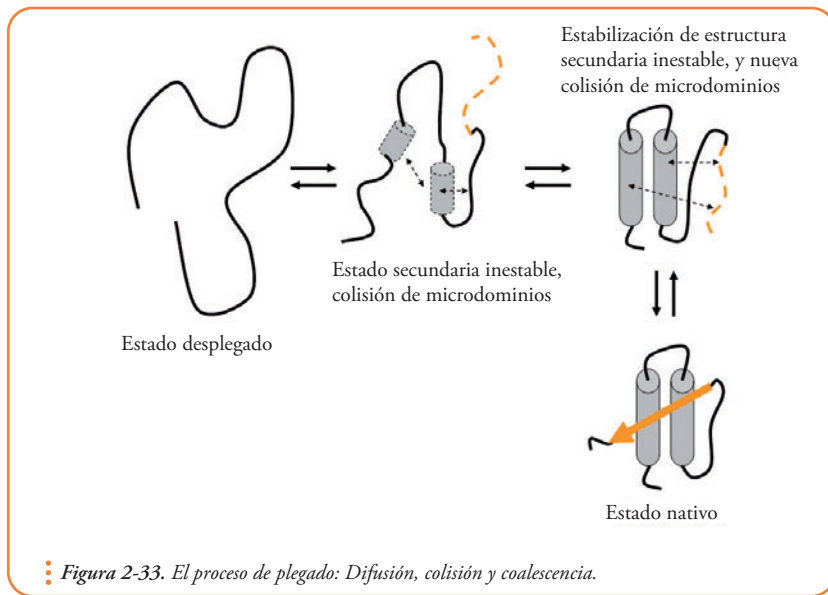
Otros modelos simplificaron la búsqueda conformacional desacoplando la formación de estructura secundaria y terciaria. Así Martín Karplus y David L. Weaver (1976) propusieron el modelo de difusión, colisión y coalescencia. Este modelo asume que las proteínas están compuestas por una serie de partes fluctuantes muy móviles (microdominios) suficientemente pequeñas y simples como para que en ellas la búsqueda conformacional sea rápida (Figura 2-33). Los microdominios no tienen acceso a estructura secundaria estable por búsqueda al azar pero sí a estructura secundaria incipiente y a grupos o clusters de interacciones hidrofóbicas.

El proceso de plegado es descrito como una serie de pasos únicos, de difusión, colisión y coalescencia, en la que se estabilizan mutuamente los microdominios. Las fuerzas que gobiernan el movimiento de microdominios son las interacciones específicas entre microdominios y las colisiones al azar con el solvente y con otras partes de la cadena polipeptídica. La estabilización final de estructura secundaria está ligada a la formación de estructura terciaria.

En la Figura las flechas discontinuas indican las colisiones de microdominios que llevarán a la estabilización de estructura secundaria. Los elementos de estructura secundaria aún inestables o insipientes se dibujaron también en trazo discontinuo.

El modelo permite postular múltiples vías (paralelas) de secuencias de pasos de difusión y colisión. La información contenida en la secuencia de cada microdominio

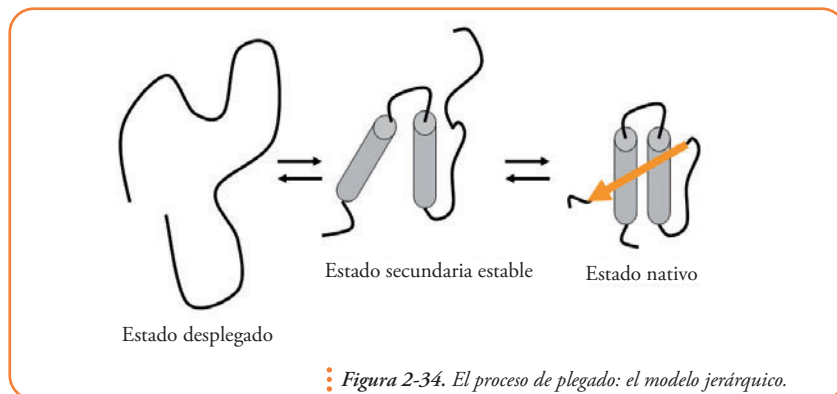
puede ser usada independientemente. Otra particularidad del modelo es que las colisiones pueden generar intermediarios con interacciones entre microdominios no adyacentes en relación a la secuencia lineal de la proteína. El modelo reduce en definitiva la dinámica del proceso de plegado, ya no considerando cada residuo, sino el análisis de las propiedades de los microdominios y sus interacciones. Este modelo predice la dependencia de la velocidad de plegado con respecto a la viscosidad del solvente.



El modelo jerárquico

En la década del 80 Robert L. Baldwin y colaboradores propusieron una serie de ideas inspiradas en los experimentos realizados por los pioneros Ptitsin, Rashin y Richmond, Richards y Cohen. Se trata de un modelo jerárquico en el que, inicialmente, se forman elementos locales de estructura secundaria nativa, independientes e intrínsecamente estables (con estructura de puentes de hidrógeno nativa), en ausencia de estructura terciaria. Posteriormente, estos elementos se empaquetan para formar estructuras más complejas (Figura 2-34).

Este modelo se ve apoyado por la evidencia de que una hélice α se forma, en muchos péptidos y proteínas, en microsegundos, mientras que la estructura terciaria tarda en aparecer milisegundos o más.



Existen numerosos ejemplos en los que la estructura terciaria es desensamblada antes que la estructura secundaria. De los experimentos en equilibrio de desplegamiento inducido por agentes caotrópicos.

El estudio de las vías del plegado (ej: $U \leftrightarrow I_1 \leftrightarrow I_2 \leftrightarrow \dots I_n \leftrightarrow N$) se hace muy complicado o imposible para proteínas pequeñas con un único dominio, porque generalmente no se observan intermediarios. Su ausencia, es explicada en el modelo jerárquico por la existencia de una alta cooperatividad en el proceso de plegado, que hace muy poco estables a los intermediarios y aproxima el proceso a un sistema de dos estados (N y U). Estos sistemas pueden abordarse con experimentos de cinética del plegado en solventes en los que los intermediarios se estabilizan.

Rompecabezas

En el año 1985, Harrison y Durbin [39] propusieron un modelo cualitativo para el mecanismo de plegado. Por analogía con el juego del rompecabezas, recibió el nombre de Rompecabezas. Los autores sugirieron que las proteínas se pliegan por vías alternativas paralelas y no por una secuencia de eventos definida. Incluso para proteínas pequeñas con unos pocos microdominios, estos, por combinatoria, podrían generar un importante número de vías alternativas para alcanzar la estructura nativa. Este modelo podría representar un caso límite de la teoría de difusión y colisiones y su aplicación podría ser dependiente de características específicas de ciertas proteínas en relación a las propiedades de los microdominios y sus interacciones (Figura 2-35).

Llevado al extremo, sería equivalente a armar el ‘rompecabezas’ comenzando por cualquier pieza. Se generarían así núcleos de plegado y la estructura crecería entorno a estos centros, pero cada molécula elegiría su propia vía hacia el estado nativo.

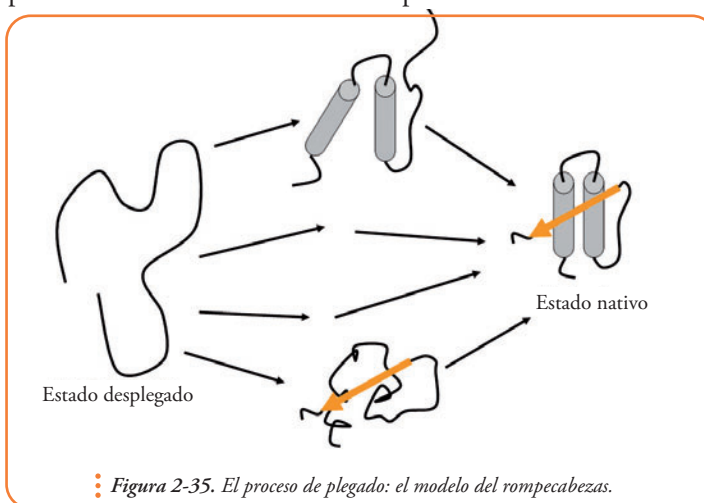


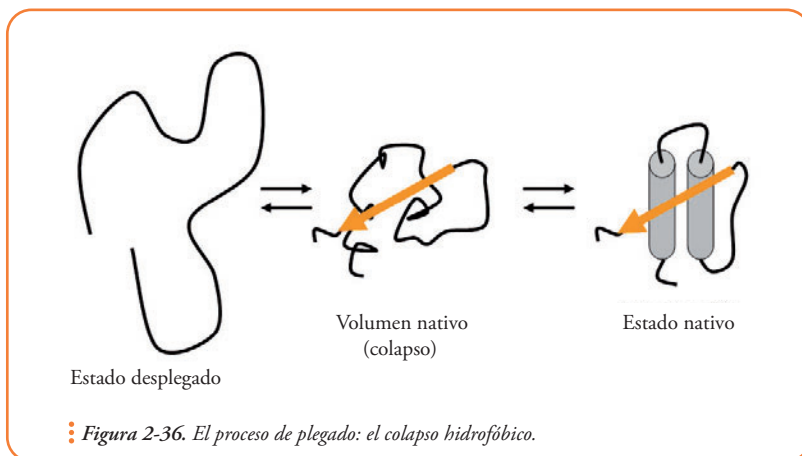
Figura 2-35. El proceso de plegado: el modelo del rompecabezas.

El colapso hidrofóbico

A fines de la década del 70 surgió la hipótesis de que un ‘colapso hidrofóbico’ podría tener lugar antes de la formación de estructura secundaria y terciaria. Bajo condiciones de replegado, el polipéptido adoptaría conformaciones compactas, incrementando notablemente la probabilidad de alcanzar la topología nativa. El efecto hidrofóbico es, para este modelo, la fuerza impulsora del plegado y, en definitiva, responsable de la deshidratación del corazón de la proteína, dando lugar a intermediarios sin estructura terciaria fija, con estructura secundaria parcial y con un volumen muy similar al nativo.

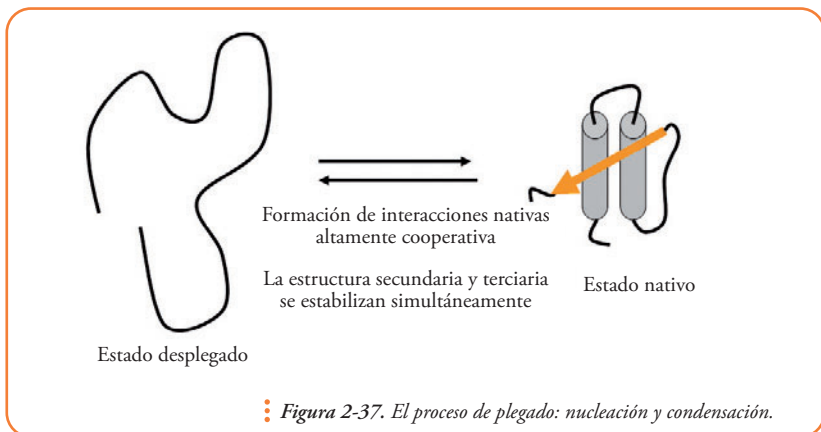
Los elementos de estructura secundaria más inestables podrían estabilizarse posteriormente. Dicho de otra forma, la formación de estructura secundaria nativa (fija) podría estar dirigida por

interacciones terciarias, o superficies apolares más o menos inespecíficas existentes en la estructura del intermediario, incluso por la topología global que adopta la cadena polipeptídica restringiendo el espacio conformacional. Hay indicios de que el colapso y la aparición de estructura secundaria podrían darse de manera simultánea (Figura 2-36).



Nucleación condensación

En el mecanismo de nucleation condensation o nucleation collapse un grupo de residuos adyacentes forma un núcleo, pero a diferencia de la visión clásica de 'nucleación', en la que el núcleo es fuertemente localizado, por ejemplo dos o tres vueltas de hélice, este modelo predice un núcleo local inestable, que se estabiliza o consolida como consecuencia de interacciones de largo alcance (terciarias) y que, en definitiva, a su vez, extienden el núcleo. La formación del núcleo no es el paso limitante porque para que éste se estabilice una fracción importante de la cadena polipeptídica debe estar correctamente posicionada para proveer los contactos de largo alcance que estabilizan al núcleo. Debe haber entonces un número crítico de interacciones locales y de largo alcance del que depende el plegado global. La estructura secundaria y la terciaria se forman paralelamente (Figura 2-37). Este modelo podría proveer el marco teórico para cada unidad de plegamiento en proteínas complejas.



Plasticidad del mecanismo de plegado

Las visiones actualizadas de los modelos jerárquicos y de difusión y colisión se sola-

pan. Un elemento de estructura secundaria estable o parcial (microdominio elemental), debería sin duda difundir y colisionar con otro o con un conjunto de microdominios para formar después un dominio mediante un ajuste fino (empaquetamiento). Las estabildades de los elementos de estructura secundaria en forma aislada podrían ser muy variables. De esta forma ambos modelos coinciden en que el plegado es un proceso jerárquico en el que se forman primero estructuras simples con estabilidad marginal, dada por interacciones locales en la secuencia, que posteriormente se agrupan para formar estructuras más complejas.

¿Existe un único mecanismo de plegado?

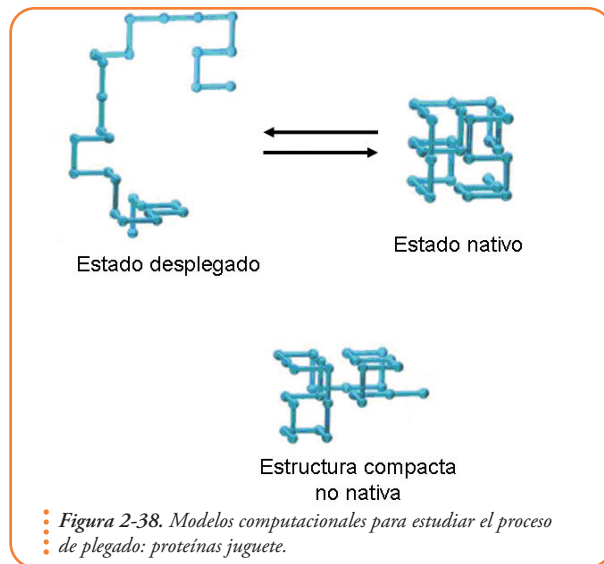
Esto parece poco probable teniendo en cuenta las evidencias experimentales. Que el plegado sea un proceso concertado, o en pasos sucesivos, podría depender de la estabilidad de las subestructuras individuales, por ejemplo, si son inestables en sus formas aisladas, la estructura final se forma de manera concertada, nucleación y condensación. Si las subestructuras individuales fueran suficientemente estables, en forma independiente, podrían alcanzar el estado nativo siguiendo un modelo jerárquico o de difusión y colisión. Las proteínas estudiadas no siguen un único modelo y aún hoy cada modelo mecanístico tiene sus adeptos y ninguno se impone claramente sobre los otros.

Paisajes energéticos y embudos de plegado

Hace algunos años se propuso una visión del plegado alternativa que permite sortear la paradoja de Levinthal y generar hipótesis con respecto al mecanismo de plegado. Esta visión surgió de la simulación extensiva en computadoras de los cambios conformacionales posibles para representaciones simplificadas de la cadena polipeptídica. En estas representaciones los residuos son considerados esferas de dos tipos, polares (P) y no polares (NP), y su ubicación en el espacio está limitada por la conectividad de la cadena y la localización en un enrejado cúbico. La energía de las interacciones entre esferas es asignada como favorable o desfavorable, de acuerdo a su carácter en polaridad. Para un número pequeño de residuos pueden ser enumeradas todas las conformaciones y su energía puede ser evaluada.

Para estas proteínas conocidas como proteínas de juguete (*toy proteins* en inglés, **Figura 2-38**) el estado nativo es aquel que presenta el mayor número de contactos favorables (NP NP, P P) y máxima compactación.

Mientras los modelos fenomenológicos anteriores usan símbolos para representar cada conformación (U o D, el estado desplegado; T₊, el estado de transición; I, estados intermedios), los nue-

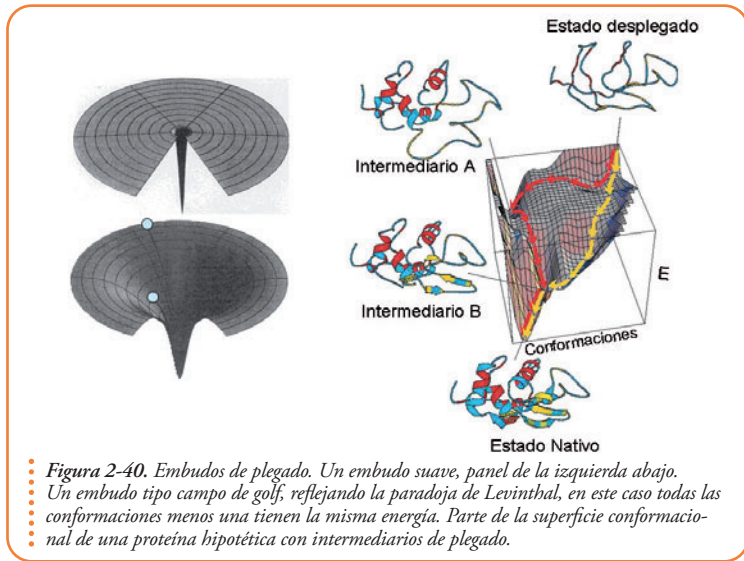
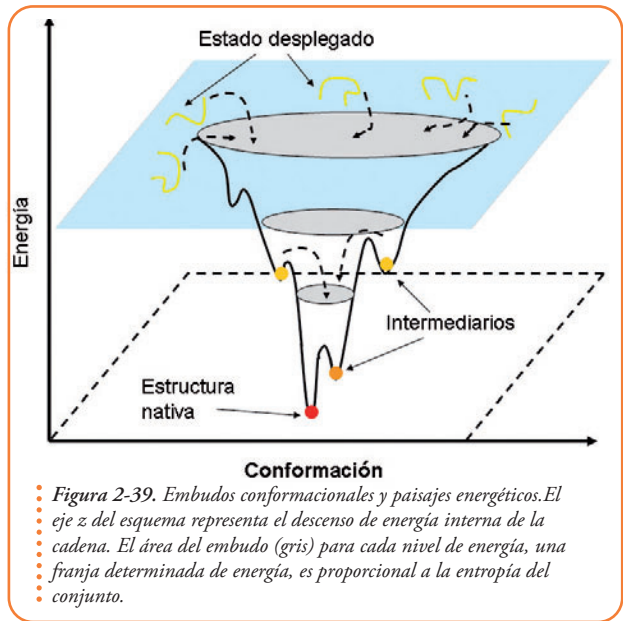


vos modelos teóricos, basados en la mecánica estadística, reconocen que esos estados macroscópicos son en realidad ‘conjuntos moleculares’ de conformaciones. Esta visión reemplaza el concepto de ‘vía de plegado’ (eventos secuenciales) con conceptos como ‘superficie conformacional’, ‘embudos de energía’ y ‘eventos paralelos’.

En estas superficies conformacionales con forma de embudos (ver **Figura 2-39**) se representa en forma gráfica el conjunto de todas las trayectorias conformacionales posibles por las que atraviesa la cadena principal desde el conjunto de estados desplegados (U) hacia la forma nativa.

Las conformaciones que son geoméricamente similares son cercanas en la superficie del embudo. El proceso de plegado puede representarse como el trazo producido por una bola al desplazarse por el embudo. La clave del modelo consiste en que cuando la energía interna y la entropía se compensa adecuadamente, el paisaje conformacional adquiere forma de embudo y esencialmente todo cambio conformacional es ‘colina abajo’, acercando la molécula al estado nativo, independientemente de la conformación inicial.

El plegado rápido, de dos estados, se corresponde con embudos suaves, sin trampas cinéticas significativas. Las cinéticas complejas, con muchas fases por el contrario, son representadas con embudos con superficies complejas, accidentadas, rugosas o con valles. En este nuevo contexto la paradoja de Levinthal deja de ser una paradoja. Llegar al mínimo global y hacerlo rápidamente dejan de ser dos aspectos mutuamente excluyentes. La paradoja es, simplemente, el producto de considerar el problema del plegamiento en términos del paisaje tipo ‘campo de golf’ (**Figura 2-40**, panel de la izquierda embudo superior), en el que la cadena busca la conformación nativa suponiendo un paisaje plano de conformaciones. En este caso todas las conformaciones están igualmente alejadas del



estado nativo en términos energéticos por lo que la chance de encontrar el estado nativo se desvanece, el paisaje ya no tiene forma de embudo. La búsqueda conformacional que propone la nueva visión del proceso de plegado es más directa, cada conformación alternativa del conjunto de las moléculas desplegadas se reconfigura rápida y progresivamente hacia el estado nativo.

Existen, entonces, procesos paralelos y conjuntos de conformaciones a los que se llega por distintas secuencias de cambios de ángulos diedros y no por ensamblado secuencial de estructuras específicas y homogéneas. En definitiva una vía está definida por una ruta unidimensional a través del espacio conformacional, mientras que los embudos, primero multidimensionales, reducen progresivamente la 'dimensión' del espacio conformacional hacia el estado nativo (prácticamente una única conformación, con cierta flexibilidad). Los intermediarios ahora pueden explicarse como mínimos de energía locales (ver Figura 2-40, panel de la derecha intermediarios A y B), valles en el embudo, o alternativamente trampas de alta entropía configuracional. En este último caso las moléculas deben gastar mucho tiempo de búsqueda (regiones llanas en el embudo). El mínimo global (conformación de menor energía teniendo en cuenta todas las conformaciones posibles es el que corresponde a la estructura nativa (ver hipótesis termodinámica de Anfinsen).

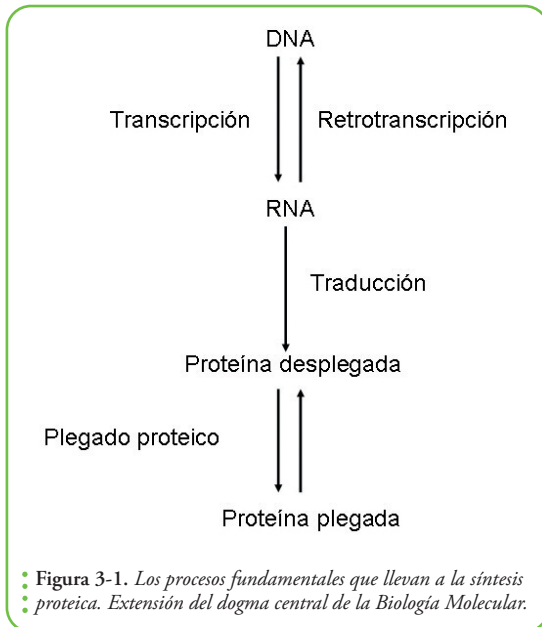
La prueba experimental del modelo requiere demostrar que el cambio de cada par de ángulos diedros no está correlacionado con el cambio de ningún otro par. Las partes superiores del embudo conformacional han sido de difícil acceso experimental porque los primeros eventos ocurren muy rápidamente. Por esto, hasta recientemente, se han estudiado las regiones más profundas, cercanas al estado nativo con barreras energéticas considerables. Sin embargo, hace muy poco tiempo comenzaron a realizarse experimentos de replegado en escala de los micro y de los nanosegundos.



Producción de proteínas en el laboratorio. Tercera Parte

El código genético. Genes. Transcripción y traducción

Los organismos vivos, como por ejemplo, nosotros o las bacterias, y los que están en el límite de lo vivo y lo no vivo, los virus, requieren para poder reproducirse o autoreplicarse una maquinaria biológica, increíblemente, compleja y organizada tanto espacial como temporalmente. La información para llevar adelante estos procesos y las instrucciones para fabricar esta maquinaria debe estar almacenada y debe poder ser transmitida a la descendencia. Para almacenarla y poder recuperarla y utilizarla es necesario que exista un código. Probablemente, uno de los más asombrosos descubrimientos haya sido la capacidad de los ácidos nucleicos, DNA y RNA de portar información. Gran parte de esa información debe ser transcrita en mensajes o mRNAs (la **m** es de mensajero) para que, posteriormente, pueda ser traducida en secuencias de proteínas por los ribosomas, que son una parte importante de la maquinaria de biosíntesis de proteínas (ver Figura 3-1).



Estas proteínas realizarán una gran variedad de funciones, por ejemplo, transportar oxígeno como en el caso de la hemoglobina, otras tendrán la función de sintetizar DNA, RNA, el caso de las DNA y RNA polimerasas y otras, por ejemplo, formarán parte de los ribosomas. Necesitamos proteínas para sintetizar proteínas.

La información genética para sintetizar proteínas está codificada, entonces, en moléculas de DNA doble cadena en la mayor parte de los organismos; aunque ciertos virus usan DNA o RNA de cadena simple o RNA de cadena doble. Las secuencias están codificadas por cuatro tipos de nucleótidos: A (adenina), T (timina), G (guanina), C (citosina) en el DNA. En el RNA la T es reemplazada por U (uracilo). Las regiones del material genético que codifican se llaman genes. Un gen puede codificar para

una o más proteínas. Pero, también, un gen puede codificar para un RNA. Existen RNAs con funciones específicas. Por ejemplo, RNAs ribosomales, RNA de transferencia y las ribozimas. Las regiones a ambos lados de cada gen, a veces incluyendo regiones muy distantes en la secuencia, las distancias, en este caso se miden en número de bases o kilobases, suelen tener funciones regu-

latorias de la expresión de esos genes. La regulación de la expresión genética es un tema apasionante y central en biología molecular.

En este punto es importante recordar que los ácidos nucleicos también son polímeros y que poseen una estructura particular. En el caso del DNA de doble cadena, la estructura más estable, por lo general, es una doble hélice. La doble hélice tiene una característica físico-química notable que soluciona entre otras cuestiones el problema de la autoreplicación de la información porque ambas hebras son complementarias. Con la información molecular contenida en una de ellas podemos reconstruir la otra y viceversa.

Cada hebra tiene un extremo 5' y un extremo 3', cada hebra entonces tiene polaridad.

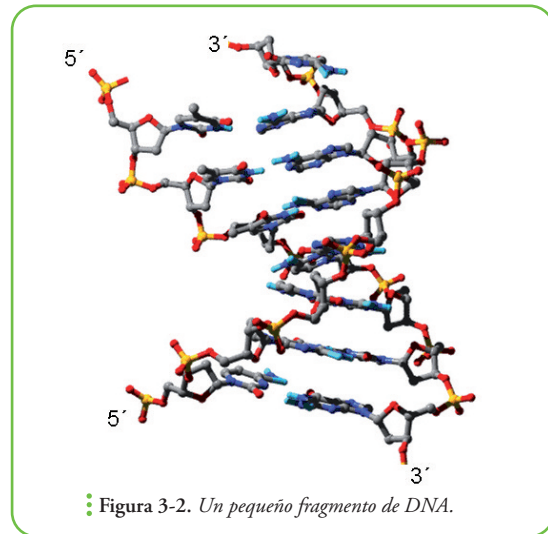
Los genes se leen en el sentido 5' a 3' y existen secuencias de iniciación y secuencias de finalización que nos permiten definir qué secuencia es un gen y qué secuencia no lo es.

A su vez, una de las hebras se define como la hebra codificante o con sentido, de esta se puede deducir qué secuencia tendrá la proteína que codifica en forma directa. La otra hebra, de polaridad opuesta *posee secuencia complementaria* se la conoce como hebra no codificante, antisentido o hebra molde.

En organismos eucarióticos, entre ellos nosotros, la parte de un gen que codifica para una proteína suele estar dividida en fragmentos llamados exones.

Los exones están separados entre sí por intrones. Los intrones, son regiones no codificantes y pueden ser muy extensos, por ejemplo 10 kilobases (10.000 bases). En cambio, en organismos procarióticos (por ejemplo las bacterias), el DNA que codifica para una cadena polipeptídica lo hace en un segmento continuo sin intrones.

Volviendo un poco para atrás, el primer paso en la biosíntesis de una proteína es la producción de moléculas mensajeras (los mRNAs). La transcripción de la secuencia de nucleótidos del DNA en un mRNA de secuencia complementaria. El descubrimiento de la existencia de los mensajeros también fue uno de esos descubrimientos importantes y maravillosos, porque en células eucariotas alguna molécula debía transportar la información codificada en el DNA, cuya ubicación es nuclear (está en el núcleo) hasta el citoplasma, donde están los ribosomas y se realiza la síntesis de proteínas. La transcripción o síntesis de un mRNA ocurre sólo si una serie de proteínas reguladoras de este proceso están correctamente posicionadas en la región del DNA que regula la transcripción del gen conocida como región promotora. Cuando esto ocurre el gen es transcrito por una

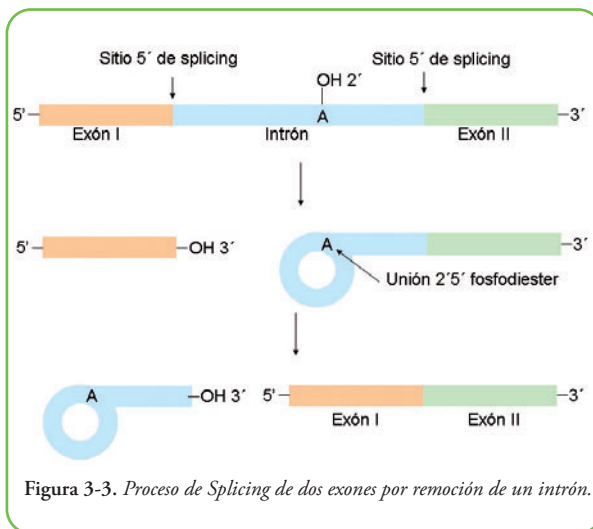


Se muestran los extremos 5' (fosfato) y 3' (-OH). Notar la estructura helicoidal del DNA doble cadena y la polaridad opuesta de ambas hebras, químicamente, complementarias. Los átomos se representan de la siguiente manera: en naranja los fosforos, en rojo los átomos de oxígeno, en azul los de nitrógeno, en gris los de carbono y en celeste los átomos de hidrógeno, que participan en puentes de hidrógeno. (ver Figura 3-2)

RNA polimerasa. También hay proteínas que pueden inhibir la transcripción, estos reciben el nombre de represores y silenciadores.

La transcripción es iniciada usando a la hebra no codificando como molde (por esto también se la llama hebra molde) es decir se sintetiza en este proceso desde 5' a 3' un polímero de RNA complementario a la hebra molde. Si comparamos este proceso en organismos procarióticos y eucarióticos, podemos ver que existen importantes diferencias. En eucariotas existen dos modificaciones del mRNA conocidas como modificaciones post transcripcionales: 5' cap y poliadenilación del extremo 3', se les agraga una cola de poliA, muchas adeninas. Ambas modificaciones son importantes porque le otorgan al mRNA mayor estabilidad y entonces es más difícilmente degradable. Además,

en el caso de los RNAs eucariotas los intrones deben ser removidos. El proceso de remoción de intrones se llama *splicing* (ver Figura 3-3).



El extremo 5' del intrón reacciona con el oxidrilo 2' de un nucleótido de adenina propio del intrón. El grupo oxidrilo 3' del exon I reaccionará atacando la unión fosfodiéster existente entre el intrón y el exón II para generar un nuevo enlace fosfodiéster con el fosfato 5' del exón II.

de los mRNA eucariotas ya sin intrones, son traducidos en cadenas polipeptídicas en el citoplasma por una maquinaria, espectacularmente, compleja y gigantesca formada por tRNAs (RNA de transferencia) rRNAs (RNA ribosomales) y una gran cantidad de proteínas. La estructura de esta maquinaria, los ribosomas y el mecanismo de biosíntesis se conoce con gran detalle.

Hasta hace relativamente poco tiempo, parecía prácticamente imposible pensar en obtener estructuras con alta resolución de los ribosomas completos.

Hoy en día, es factible, para cualquiera de los lectores de este párrafo, que cuente con una computadora conectada a Internet, obtener modelos experimentales de alta resolución de las estructuras atómicas de ambas subunidades ribosomales.

Inclusive, se pueden obtener de la PDB estructuras en las que se han incorporado al mRNA y tRNA,

Una maquinaria muy compleja llamada spliceosoma es la que se encarga de hacerlo. Algunas variaciones en este proceso conocidas como *splicing* alternativo combinan a los exones de manera alternativa, por ejemplo eliminando uno. Por lo que un único gen puede codificar, en realidad, una gran variedad de proteínas. Se cree que en el ser humano cerca del 50 % de los genes son afectados por algún tipo de *splicing* alternativo.

Este mecanismo contribuye de manera muy importante con la diversidad de proteínas existentes en el interior de una célula. Algunas de ellas pueden tener diferencias funcionales importantes.

Los mRNA, maduros en el caso

La traducción (biosíntesis ribosomal de proteínas) puede reproducirse, experimentalmente, en sistemas simplificados a partir de extractos libres de células obtenidos por la ruptura mecánica de las células y separación y/o purificación de los componentes necesarios que forman parte de la maquinaria biosintética, precursores y la fuente de energía necesaria. La traducción del mRNA en proteínas es realizada mediante *la lectura* (no hay lector sino moléculas que interactúan entre sí reconociéndose) del código genético: cada triplete o codón de la región codificante de nucleótidos de un mRNA tiene su significado: información para un aminoácido o para terminar la síntesis de la proteína. Cada uno de los cuatro nucleótidos (A, G, U, o C) puede ocupar cada una de las tres posiciones en un codón. Así, hay $4^3=64$ posibles combinaciones de secuencias de tres nucleótidos. Se ve, fácilmente, que hay más combinaciones que los 20 aminoácidos con los que se sintetizan las proteínas. La asignación sistemática de cada codón a un aminoácido o a una señal de terminación (**STOP**) de la traducción fue un paso importantísimo en el campo de la biología molecular y permitió dilucidar por completo el código genético (Figura 3-4).

Con esta asignación y la secuenciación de DNA se hizo posible determinar cuál es la secuencia de aminoácidos que codifica un gen directamente a partir de la secuencia de nucleótidos. Entonces la secuencia de DNA en sentido 5' hacia 3' pero leída en tripletes corresponde a la secuencia de aminoácidos de la proteína codificada leída en dirección desde el N-terminal hacia el C-terminal.

Una característica notable del código genético es que se trata de un código degenerado. La mayor parte de los aminoácidos pueden ser codificados por más de un codón. Salvo metionina y triptofano, que son las únicas dos excepciones, codificados únicamente por un triplete, AUG y UGG, respectivamente. Los codones que codifican para el mismo aminoácido se llaman codones sinónimos.

		SEGUNDO NUCLEÓTIDO				
		U	C	A	G	
PRIMER NUCLEÓTIDO	U	Phe	Ser	Tyr	Cys	U
		Phe	Ser	Tyr	Cys	C
		Leu	Ser	STOP	STOP	A
		Leu	Ser	STOP	Trp	G
	C	Leu	Pro	His	Arg	U
		Leu	Pro	His	Arg	C
		Leu	Pro	Gln	Arg	A
		Leu	Pro	Gln	Arg	G
	A	Ile	Thr	Asn	Ser	U
		Ile	Thr	Asn	Ser	C
		Ile	Thr	Lys	Arg	A
		Met	Thr	Lys	Arg	G
	G	Val	Val	Asp	Gly	U
		Val	Val	Asp	Gly	C
		Val	Val	Glu	Gly	A
		Val	Val	Glu	Gly	G

Figura 3-4. El código genético. Cada aminoácido está codificado por uno o más codones (grupos de 3 bases).

Un punto importante, particularmente para los ingenieros y productores de proteínas, es que existen frecuencias de uso de codones sinónimos.

Por ejemplo, si empleamos un organismo procarionta para producir en forma recombinante una proteína eucariota puede ocurrir que alguno de los codones sea muy poco utilizado y la síntesis de proteínas a partir de mRNAs que contengan este tipo de codones sea ineficiente. Como resultado produciríamos poca proteína. Los codones que representan el mismo o un aminoácido con propiedades fisicoquímicas relacionadas, suelen tener secuencias similares. En muchos casos la tercera base del codón suele no ser significativa: los cuatro codones, difiriendo sólo en la tercera base, pueden representar al mismo aminoácido. En otros casos, queda determinada la asignación del aminoácido si

se trata de una purina (A o G) o una pirimidina (U o C) como tercera base del codón, AGU y AGC codifican para el aminoácido serina; AGA y AGG para arginina. La reducción de especificidad en la última posición es conocida como degeneración de la tercera base.

Por otro lado hay que resaltar que 3 de las 64 posibilidades que corresponden a los codones UAA, UAG y UGA no representan aminoácidos sino señales de terminación de la traducción o señales *stop*. Estos codones especifican la finalización de la síntesis de la proteína que está siendo sintetizada.

La síntesis de proteínas se basa en el ensamblado de una línea de producción en la que los ribosomas proceden con un elevado consumo de energía (provista por la hidrólisis de GTP a lo largo del mensajero) trayendo los aminoacil-tRNAs que proveerán los bloques de construcción del producto proteico.

El proceso es complejo y están involucradas muchas reacciones químicas. La iniciación, que es la **primera etapa**, involucra reacciones que preceden la formación del enlace peptídico entre el primer y el segundo aminoácido de la proteína. Esta etapa requiere que el ribosoma se una al mRNA para formar un complejo, conocido como complejo de iniciación que contiene al primer aminoacil-tRNA. Este paso, por lo general, determina la velocidad de la traducción porque es un paso lento. La secuencia en el mRNA en la que ocurre la iniciación en *Escherichia coli* es el sitio de unión del ribosoma (en inglés RBS) o Shine-Dalgarno; es una secuencia corta (5' AGGAGG 3') que precede al AUG en menos de 10 bases (AAACAGGAGGAUUACCCAUG). Esta distancia es muy importante para que la iniciación tenga lugar y **debe** tenerse muy en cuenta al diseñar construcciones genéticas para producción de proteínas recombinantes en bacterias. La secuencia RBS es reconocida por la subunidad menor del ribosoma mediante la interacción con un rRNA (RNA ribosomal 16S) que contiene la secuencia complementaria 3' UCCUCC 5' al mRNA formando el complejo de iniciación, entonces se une la subunidad mayor generando el ribosoma completo.

La **segunda etapa** es la elongación. Esta etapa incluye todas las reacciones desde la síntesis del primer enlace peptídico entre el primer aminoácido que siempre es una metionina y el segundo hasta la adición del último aminoácido.

Los aminoácidos se van agregando de a uno a la cadena polipeptídica que se va sintetizando. La adición de un aminoácido a la cadena es un paso rápido en la síntesis proteica. El aminoacil-tRNA entrante, que aportará el próximo aminoácido, ocupa una región del ribosoma conocida como sitio A. Este sitio expone el codón del mRNA complementario al anticodón del tRNA. Por otro lado en otro sitio, conocido como sitio P, está el peptidil-tRNA (la proteína que se está sintetizando). El enlace peptídico se forma como consecuencia de una reacción en la cual el polipéptido que es llevado como peptidil-tRNA es transferido al aminoacil-tRNA que está posicionado en el sitio A. Esta transferencia es catalizada por componentes de la subunidad ribosomal mayor.

Así, un nuevo peptidil-tRNA fue creado en el sitio A. Este peptidil-tRNA es más largo (un aminoácido más largo) que el peptidil-tRNA existente en el ciclo anterior. El tRNA, desde el que se transfiere el peptidil luego de esta reacción de transferencia queda sin peptidil (deacilado) y ya no es afín por el sitio P pero deja el ribosoma desde un sitio transitorio llamado sitio E, pero mientras este t-RNA deacilado está en el sitio

E la afinidad del sitio A por un nuevo aminoacil-tRNA es muy baja, esto no permite que otro aminoacil-tRNA ingrese hasta que el ribosoma esté listo.

Luego se produce un movimiento molecular por el cual el ribosoma queda posicionado en el siguiente codón. Este proceso es la translocación. Con este movimiento el nuevo peptidil-tRNA queda ubicado en el sitio P y el sitio A expone el siguiente codón vacante y con afinidad incrementada para la entrada del siguiente aminoacil-tRNA. Así se repite el ciclo.

Por último, la **tercera etapa** involucra la liberación del polipéptido que se sintetizó completo y, al mismo tiempo, la disociación del complejo ribosoma-mensajero para dar ribosoma libre más mensajero libre. La biosíntesis ribosomal de proteínas es un proceso relativamente rápido y depende en gran medida de la temperatura. Por ejemplo, a temperaturas bajas (10°C) es muy lento. En bacterias a 37°C se agregan unos 15 aminoácidos por segundo al polipéptido. La dependencia de la velocidad de síntesis con de la temperatura es otra variable que los proteinólogos tienen siempre en cuenta a la hora de producir una proteína de interés. Por ejemplo, a veces, se obtienen buenos rendimientos cuando se disminuye la velocidad de síntesis, cultivando las bacterias a 30°C, en este caso los cultivos deben crecerse por más tiempo.

Producción de proteínas recombinantes en bacterias

Para hacer los estudios estructurales básicos de una proteína de interés son necesarios, aproximadamente, unos 10 ó 20 mg de proteína pura. Esto, que parece un problema sencillo, ha sido en el pasado uno de los problemas centrales en la biología estructural. La mayor parte de las proteínas se produce en muy bajas concentraciones dentro de la célula. Por otro lado dentro de una célula coexisten miles de proteínas diferentes, cerca de unas 10000 proteínas diferentes.

Probablemente, una de las cuestiones más intrigantes es:

¿Cómo podemos hacer para producir una proteína?

La respuesta hoy en día no es tan difícil de responder. Sin embargo, hace unos 60 años podría haberse tratado de un capítulo de ciencia ficción.

La clave está en que podemos usar a otros organismos como fábricas. En este ejemplo vamos a usar una cepa de bacterias de la conocidísima especie *Escherichia coli* y vamos a ver que se puede cultivar a estas bacterias y controlarlas, de tal manera, que produzcan la proteína de interés. Pero para producir proteínas en *E. coli* necesitamos saber algunas cosas.

Tecnología del DNA recombinante.

Como hemos visto, la maquinaria de biosíntesis de proteínas de cualquier organismo funciona reconociendo señales que están codificadas en los ácidos nucleicos en el DNA y en el RNA. Así que *para poder pedirle a estas bacterias algunos favores* tendremos que entender muy bien el lenguaje molecular. Concretamente, el favor será que después de haber incorporado un gen foráneo utilice la información codificada en la secuencia del gen para la síntesis de una proteína de interés. La tecnología para llevar a delante este proceso hasta la biosíntesis de la proteína se llama tecnología del DNA recombinante; combinaremos DNA

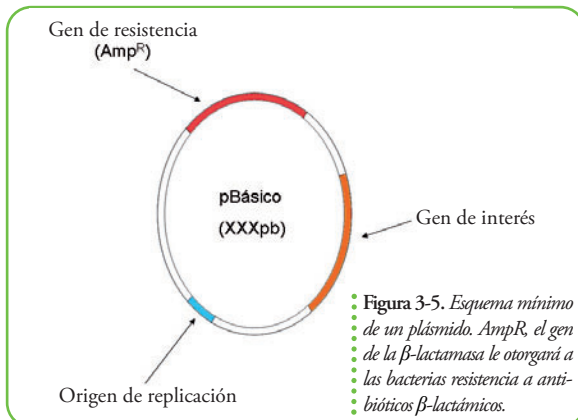
de origen bacteriano con DNA sintetizado en el laboratorio. Cualquier error en nuestras construcciones moleculares o en el dictado de las instrucciones llevaría a que la maquinaria de biosíntesis no lleve adelante el proceso biosintético que nosotros deseamos que realice.

Por esto, como primera medida debemos conocer la estructura mínima que debe tener el mensaje para ser realmente interpretado como mensaje.

El mensaje, como vimos, recibe el nombre de gen. Si bien no todos los genes codifican mensajes para la síntesis de proteínas, aquí nos enfocaremos en entender la estructura de un gen destinado a la síntesis de proteínas en un sistema de producción sencillo y comúnmente usado.

El gen de interés que no es otra cosa más que un fragmento de DNA con características especiales, debe introducirse en un sistema que se replique y que se distribuya cada vez que una bacteria de nuestro cultivo se divida en las bacterias de la siguiente generación. Esto se logra insertando el gen en un sistema que se autorreplique de manera sincronizada con la replicación del cromosoma bacteriano. Los sistemas más comúnmente usados son los plásmidos. Cada plásmido está formado por una molécula de DNA doble cadena circular y consta de un conjunto pequeño de genes propios del plásmido. Para replicarse dentro de la bacteria los plásmidos deben tener un sitio llamado origen de la replicación que debe ser reconocido por la maquinaria de replicación de DNA de la bacteria en la que se replica, en este caso *E. coli*. En la mayor parte de los plásmidos tiene algún gen que le otorga a la bacteria en la que ingresa alguna ventaja con respecto a las bacterias que no lo tienen.

Típicamente, esta ventaja es un gen para la síntesis de alguna proteína que le da resistencia contra un antibiótico presente en el medio en el que se las cultiva, en nuestro ejemplo es la resistencia a la ampicilina, esta resistencia estará dada por la presencia del gen de la β -lactamasa. Esto garantiza que todas las bacterias del cultivo contengan en su interior al plásmido. Las bacterias además de sintetizar la proteína de interés sintetizarán entre otras macromoléculas a la β -lactamasa.



En la **Figura 3-5** se muestra un esquema mínimo de la organización de un plásmido.

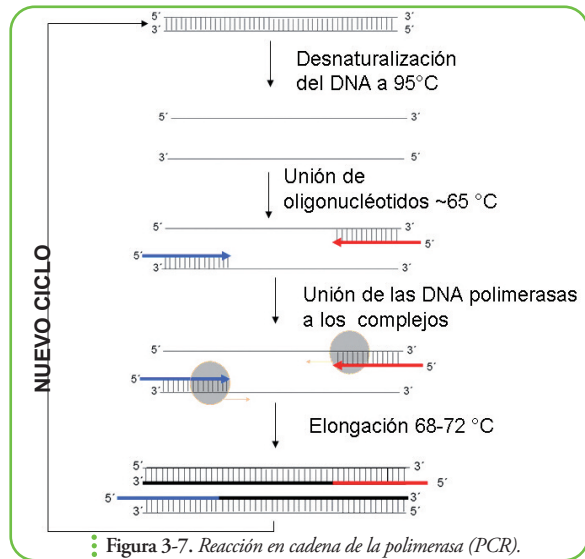
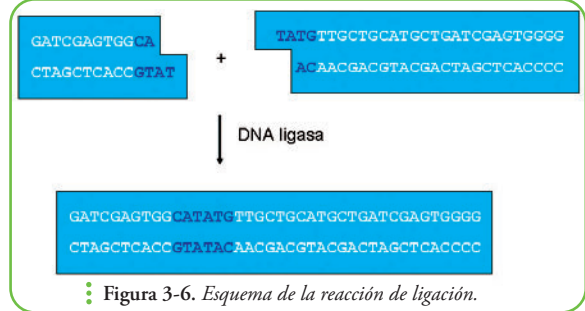
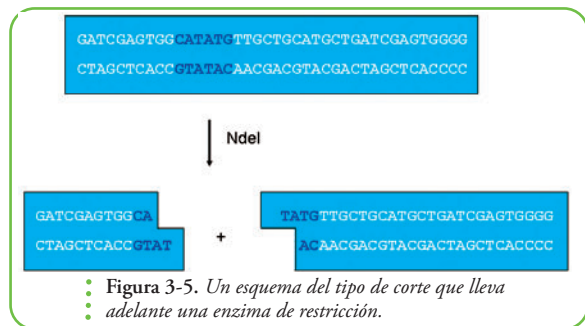
Por suerte contamos con algunas herramientas moleculares que llevaron muchos años de descubrimientos y de desarrollo porque, para poder insertar genes de interés en plásmidos, vamos a necesitar herramientas de diversos tipos:

- **Tijeras moleculares de DNA:** necesitaremos enzimas que catalicen el corte de la doble cadena de DNA en sitios extremadamente específicos. Estas enzimas se conocen con el nombre genérico de **enzimas de restricción** y las más comunes reconocen secuencias palindrómicas en el DNA, por ejemplo Nde I reconoce la secuencia 5' CATATG 3' 3' GTATAC 5'

- **Pegamento:** vamos a necesitar enzimas que sean capaces de detectar los extremos de dos fragmentos de DNA y unirlos estas enzimas son las **DNA ligasas**.

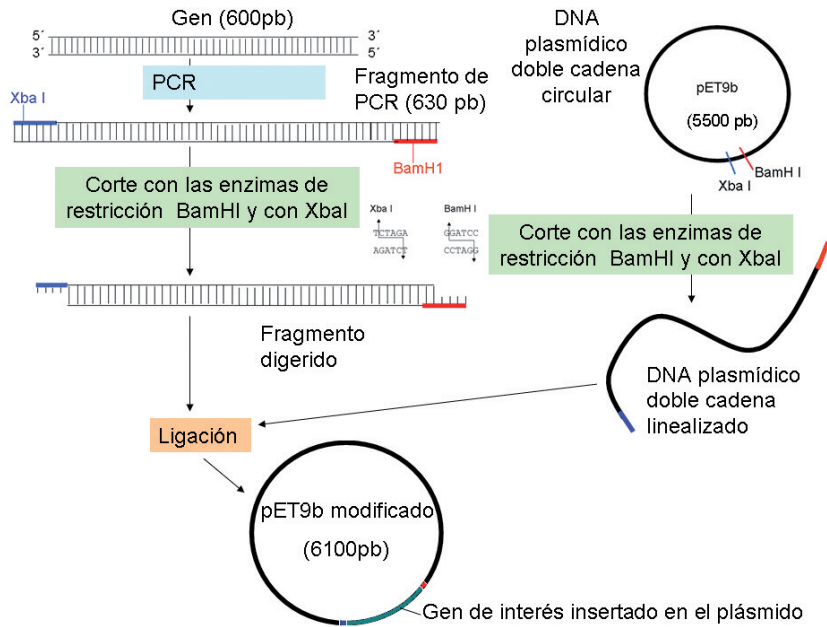
Por otro lado, vamos a necesitar producir (amplificar) muchas moléculas idénticas de DNA para poder trabajar. Las bacterias se encargarán, por lo general, de producir plásmido y, así, podremos purificar gran cantidad de moléculas de plásmido con relativa facilidad. Sin embargo, cuando queramos producir en gran cantidad fragmentos pequeños o introducir mutaciones puntuales en el sitio específico en el gen de interés, usaremos una técnica, ahora muy común, conocida con el nombre de PCR, la reacción en cadena de la polimerasa. Vale la pena resaltar que la invención de la técnica de la PCR ha marcado un punto de inflexión en la biología molecular. Se trata de una reacción en la que se utilizan pequeños oligonucleótidos (conocidos con el nombre de *primers*) sobre los que se inicia el proceso de síntesis de DNA (elongación) en el que la DNA polimerasa usa como molde una de las dos hebras de DNA generando una copia fiel de la hebra complementaria. Esta reacción se lleva adelante en varios ciclos (típicamente 30 ciclos) y el número de moléculas de DNA se eleva con cada ciclo aproximadamente en forma exponencial (ver **Figura 3-7**). La clave de este proceso fue el descubrimiento y la aplicación de **DNA polimerasas** termoestables, que resisten al proceso de desnaturalización del DNA y funcionan a temperaturas cercanas a los 70 °C.

Así por PCR, entre otras cosas, podemos introducir sitios de restricción adecuados en el gen de interés, para poder insertarlo en un sitio específico. Usando enzimas de restricción cortaremos el fragmento de PCR para dejar los extremos preparados para su ligación en el plásmido. Para poder ligar el fragmento, el plásmido deberá digerirse con enzimas que generen extremos compatibles, esto se ejemplifica en la **Figura 3-8**.



Un ciclo típico. En la reacción deben estar presentes un conjunto de reactivos; entre ellos los desoxinucleótido trifosfato (dATP, dTTP, dGTP y dCTP), iones magnesio, un buffer para la regulación de un pH óptimo, la polimerasa, los oligonucleótidos iniciadores o primers; la complementación de las hebras de DNA se marca con pequeñas rayas negras verticales, los oligonucleótidos sintéticos se señalan con una flecha azul y con una roja y se indica la polaridad de cada hebra de DNA. (ver Figura 3-7)





• **Figura 3-8.** Reacciones para el subclonado de un fragmento de DNA que codifica para un gen. Inserción de sitios de restricción por PCR y digestión con enzimas de restricción del fragmento y de un plásmido de expresión. Luego a través de una reacción de ligación quedará preparado el nuevo plásmido para ser introducido en bacterias.

Aquí, tanto el fragmento producto de la amplificación por PCR como el plásmido se cortaron con dos enzimas Xba I y BamH I esto determina que el fragmento únicamente se pueda ligar en el sentido correcto y, así, orientamos correctamente el inserto regenerando por complementación de bases del DNA los sitios de restricción.

Esto, muy comúnmente, se hace por electroporación de las bacterias. Se las somete a voltajes extremos (2500 V) durante un muy pequeño instante, esto origina micro poros por los que pueden ingresar las moléculas de DNA plasmídico.

Ahora debemos reintroducir la construcción quimérica en bacterias. Sin embargo existen algunos detalles importantes que debemos tener muy en cuenta. De lo contrario no será posible inducir la expresión de proteína de interés (proteína recombinante):

a. Los genes que se expresan en bacterias requieren regiones promotoras de la transcripción del RNA mensajero (mRNA) correspondiente. Sin un promotor adecuado no habrá síntesis de mRNA. Estas regiones promotoras serán reconocidas por las RNA polimerasas que son las enzimas encargadas de la síntesis del mRNA.

b. También debemos tener presente que los genes a su vez pueden tener regiones de silenciamiento o represión de la transcripción. En estas regiones se unen proteínas que reconocen estas secuencias específicas e impiden la transcripción porque bloquean

la unión de las RNA polimerasas o impiden su avance. Ambas regiones (promotor más sitio de unión de represores) se conocen con el nombre de región operadora u operador, o región regulatoria de la transcripción.

c. Además, el gen debe contar con una secuencia de terminación de la transcripción, estas son secuencias que causan el desensamblado del complejo mRNA:RNA polimerasa y limitan la longitud del mRNA. Suelen ser regiones ricas en bases GC. Esto determina la presencia de regiones estructuradas en el RNA.

d. Otro punto importante es que, para que se lleve adelante la traducción, debe existir un sitio de iniciación que será reconocido por el ribosoma. Unas 10 pares de bases hacia el extremo 5' del codon ATG (AUG en el mRNA) que codifica para el primer residuo de la cadena polipeptídica (una metionina) debe estar presente la secuencia RBS o sitio de unión del ribosoma, también conocida como secuencia Shine-Dalgarno: AGGAGA. Si esta secuencia no estuviera presente el ribosoma bacteriano no reconocería el sitio de iniciación y tampoco sería posible inducir la expresión de proteína de interés.

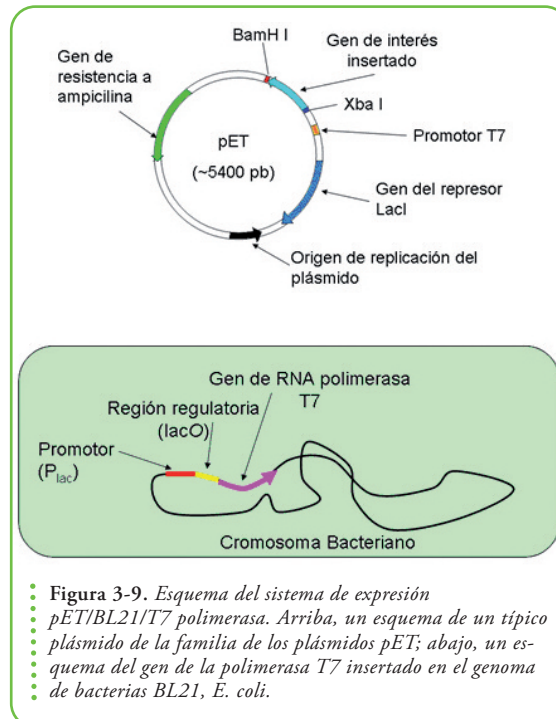
Debemos recordar que los ribosomas, que son la maquinaria que va a estar encargada de sintetizar la proteína utilizarán al RNA mensajero en la síntesis en mRNA estará codificada la información para ser traducida en una secuencia de aminoácidos.

e. Además, debe haber un codón de terminación de la traducción (*STOP*, en inglés) al finalizar la secuencia de codones que codifican para la proteína.

Hace ya unos cuantos años fue desarrollado un sistema que, actualmente, sigue siendo muy comúnmente usado para la producción de proteínas recombinantes en *E. coli*. Es un sistema de complejidad intermedia. Se trata del sistema de expresión basado en los plásmidos pET.

En este caso, se utiliza una cepa de bacterias que tiene algunas modificaciones genéticas. Entre ellas una de las más relevantes es que en el cromosoma bacteriano se introdujo el gen de una polimerasa de origen viral (del virus bacteriofago T7).

La transcripción del gen de la RNA polimerasa viral está en condiciones normales impedida, controlada por un represor; por técnicas de ingeniería genética se ha insertado en la región regulatoria del gen una secuencia que es reconocida por un represor, el represor LacI (lac de lactosa, el gen que codifica para este represor también ha sido insertado en el plásmido, ver **Figura 3-9**). Sin embargo, en presencia de lac-



tosa, o moléculas miméticas (la más común es el IPTG) el represor deja de unirse al sitio de regulación (operador) en el DNA y se desreprime la transcripción del gen de la RNA polimerasa T7.

Dicho de otra manera, el gen de la RNA polimerasa del bacteriofago se transcribe, fuertemente, en presencia de lactosa o IPTG generando gran cantidad de copias del mRNA. Este mRNA se traducirá mediante la maquinaria de traducción bacteriana, generando una elevada concentración de la proteína RNA polimerasa del fago.

Un esquema de algunos de estos aspectos se muestra en la **Figura 3-10**.

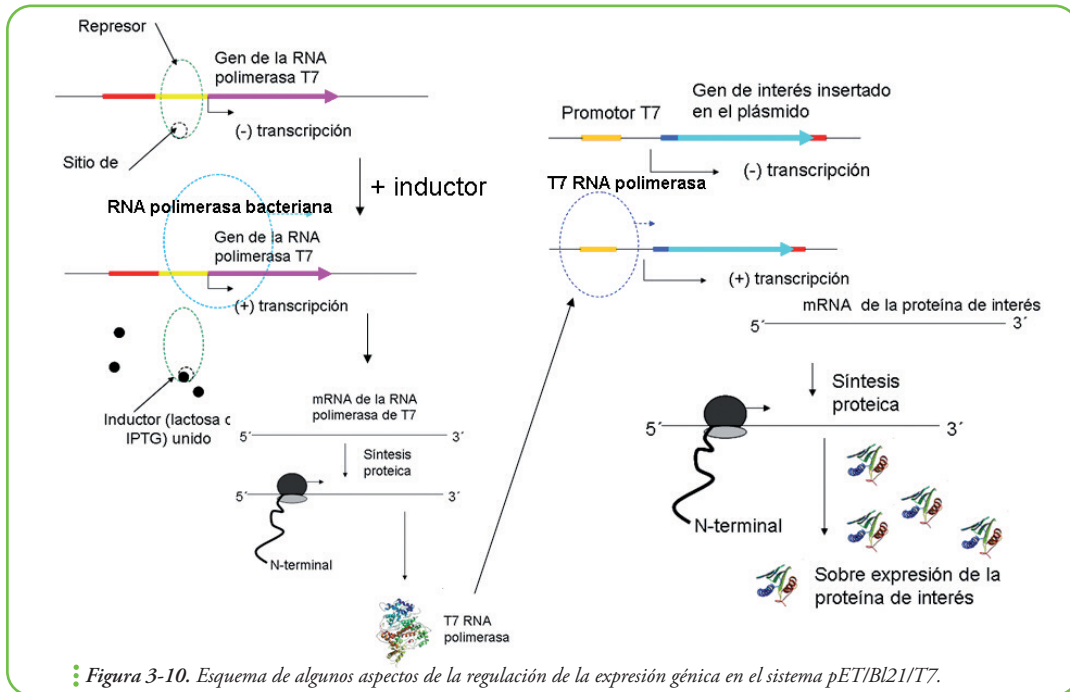


Figura 3-10. Esquema de algunos aspectos de la regulación de la expresión génica en el sistema pET/BL21/T7.

Es decir que el gen únicamente va a ser transcrito por esta RNA polimerasa y no por las RNA polimerasas endógenas de la bacteria, por esto se dice que es *inducible*. En unas pocas horas, después de haber agregado el inductor IPTG, se habrán transcrito suficientes moléculas del mRNA correspondiente a la proteína de interés y, a partir de este, se habrá traducido el mensaje dando como resultado la expresión (producción) de miles de moléculas de proteína.

Para aumentar la eficiencia de la producción se utilizan medios de cultivo ricos en fuentes de carbono y nitrógeno. Además se asegura la presencia de todos los minerales necesarios para el crecimiento adecuado de las bacterias. Los cultivos se realizan en medios líquidos y se garantiza una excelente transferencia de oxígeno requerido para el crecimiento aeróbico con una fuerte agitación de los frascos (típicamente se usan unos

Lo importante es que el gen de interés insertado, adecuadamente, en el plásmido queda bajo el control de una región promotora específica para la unión de la RNA polimerasa del bacteriofago T7 (promotor T7).

frascos especiales conocidos con el nombre de erlenmeyers). Después de unas tres o cuatro horas de cultivo con agitación a 37° C, podremos cosechar las bacterias. La forma más común de hacerlo es la centrifugación.

Para poder separar las bacterias del medio de cultivo usaremos una centrífuga. La centrifugación es un método por el que pueden separarse partículas presentes en líquidos. El aparato que hace este proceso es la centrífuga. La centrífuga aplica a la mezcla un movimiento circular; que provoca la sedimentación del sólido o de las partículas de mayor densidad. Para poder centrifugar las bacterias habrá que transvasar el cultivo a otro tipo de recipientes resistentes a las fuerzas centrífugas. A medida que las partículas que, en nuestro caso, son las bacterias comiencen a percibir la aceleración, producto del movimiento circular de un rotor (unas 5000 revoluciones por minuto (rpm) por 15 minutos serán suficientes), se irán depositando en el fondo de estos recipientes. Finalmente, si eliminamos el medio líquido remanente obtendremos una especie de pasta llamada *pellet*, o material insoluble. El *pellet*, en este caso, está formado efectivamente por bacterias que en su interior contienen la proteína de interés. Por litro de cultivo de bacterias, en condiciones normales, puede obtenerse entre 3 y 5 g de *bacterias inducidas*.

Mediante geles de poliacrilamida en presencia de SDS (ver Capítulo IV) podremos verificar, fácilmente, la expresión (producción) de la proteína.

Para esto tenemos que romper las bacterias y evaluar si la proteína de interés se conserva en forma soluble.

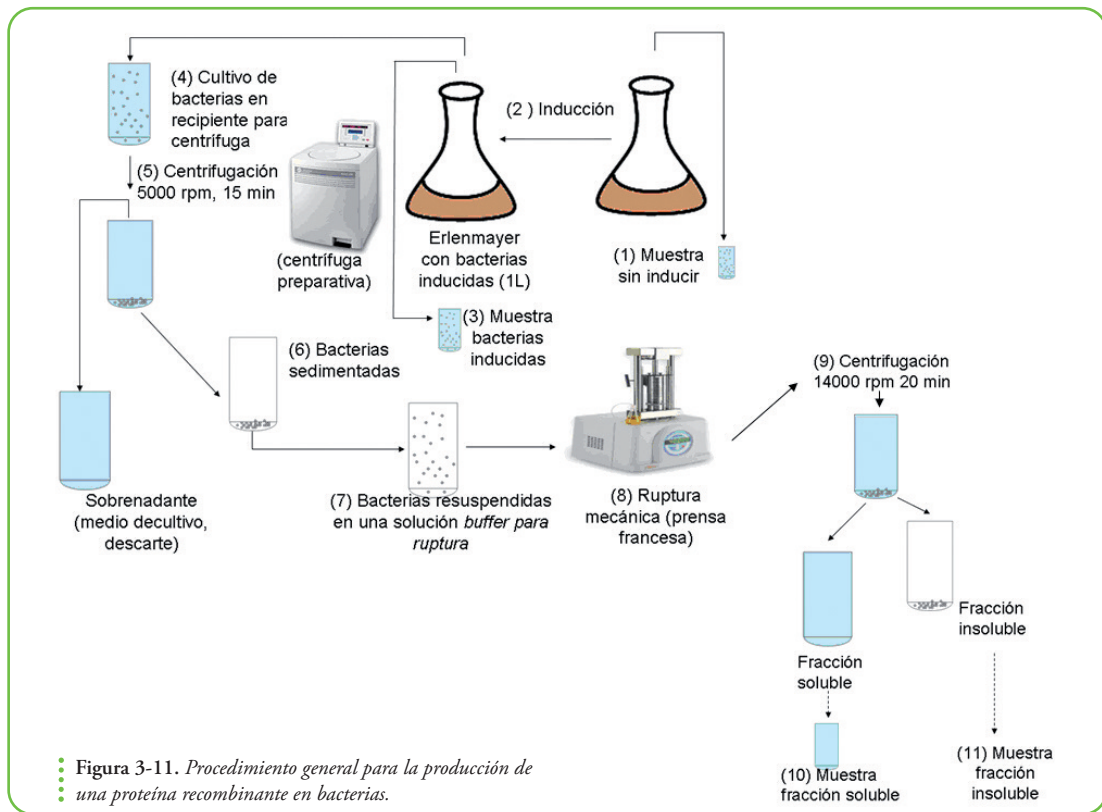
Hay diversas maneras de romper las bacterias:

- a. con cambios bruscos de presión y descompresión (el equipo se conoce como prensa francesa);
- b. con ultrasonido (con un sonicador).

Cuando rompemos las bacterias, nuevamente, tenemos que separar los fragmentos de bacterias y las proteínas insolubles de las proteínas solubles. Para esto, recurrimos a la centrifugación, esta vez deberemos centrifugar a mayor velocidad unas 14000 rpm por 20 minutos). Después de centrifugar podremos separar una fracción soluble, el *sobrenadante* con el cuidado de no resuspender el material insoluble. En un gel de poliacrilamida podremos sembrar muestras de estas fracciones e identificar a la proteína de interés y, finalmente, evaluar si se encontraba o no en la fracción soluble. En la **Figura 3-11** se muestra un diagrama del procedimiento general.

Será muy importante tomar una muestra del cultivo antes de agregar el inductor de la expresión, IPTG 1 mM, concentración final en nuestro caso. Esta muestra de bacterias sin inducir permitirá identificar por comparación con las bacterias en las que se agregó inductor cuál es la proteína de interés y cual es su nivel de expresión.

Será muy importante evaluar si la proteína de interés se encuentra soluble en el interior de las bacterias o no.



Por supuesto aún falta uno de los procesos más importantes y más complicados para poder contar con unos 10 ó 20 miligramos de proteína con que hacer estudios sobre su estructura. Debemos purificarla del resto de las otras proteínas y macromoléculas. Típicamente, después de la ruptura, ambas fracciones soluble e insoluble suelen quedar contaminadas con ácidos nucleicos y la fracción insoluble suele contener lípidos, fosfolípidos y fragmentos de pared bacteriana.

Será muy importante haber definido, correctamente, la ubicación de la proteína de interés porque dependiendo de esto, si se ha mantenido soluble o no, las estrategias de purificación serán completamente distintas e, incluso, si se halla en la fracción insoluble tendremos que intentar replegarla. Como vemos, nuevamente, la biosíntesis de la cadena polipeptídica no garantiza que la conformación de ésta sea la correcta.

¿Por qué la proteína de interés puede hallarse en la fracción insoluble?

Durante la sobre expresión, tanto la velocidad de síntesis (de la traducción) como la cantidad de moléculas producidas, son parámetros que pueden afectar, profundamente, la solubilidad de la proteína de interés. En muchos casos las proteínas no logran plegarse correctamente e interaccionan entre sí (se agregan) formando precipitados intracelulares insolubles que se conocen como cuerpos de inclusión.

En estos casos nuestra tarea será, primero, purificar los cuerpos de inclusión y, luego, di-

solverlos con un agente desnaturizante (urea, por ejemplo), desarmando los agregados y llevando a las moléculas a la conformación desplegada.

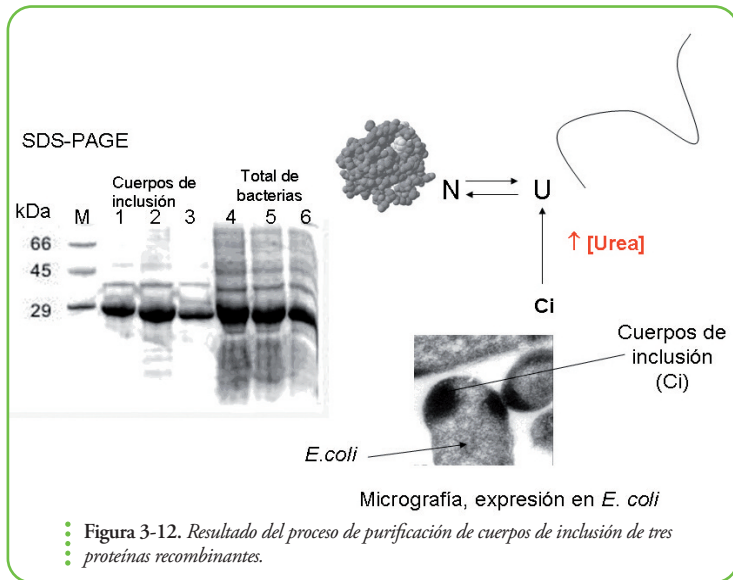
Como los cuerpos de inclusión suelen estar formados, básicamente, por la proteína recombinante más algunos pocos contaminantes, como fragmentos de pared bacteriana (peptidoglicano), lípidos y DNA, después de una serie de tratamientos quedarán relativamente enriquecidos en agregados de la proteína recombinante. Los tratamientos consistirán en lavados con soluciones que contengan enzimas que degradan el DNA (nucleasas), detergentes que permiten solubilizar los lípidos (sin solubilizar la proteína de interés) y lisozima, de clara de huevo que permite degradar los fragmentos de pared bacteriana.

En la **Figura 3-12**, se muestra el resultado de la purificación de cuerpos de inclusión de tres proteínas recombinantes (calles 1, 2 y 3 del gel). También, para poder comparar se sembraron muestras de bacterias después de su ruptura pero sin separar fracciones (calles 4, 5 y 6). Comparando la banda de interés con el resto de las bandas proteicas puede verse, además, que el nivel de sobreexpresión en este caso fue muy elevado. Además se sembraron marcadores de peso molecular (M) para poder distinguir el tamaño de las proteínas recombinantes.

Como se mencionó anteriormente, en este caso la siguiente etapa es desplegar la proteína presente en los cuerpos de inclusión por la adición de agentes desnaturizantes y llevarla al estado desplegado (U) para, posteriormente, continuar su purificación y replegado hacia la forma nativa (N).

Una de las estrategias de purificación más común es la cromatografía de intercambio iónico. La cromatografía de intercambio iónico permitirá separar proteínas de acuerdo con su carga. Un cálculo, relativamente, simple nos permitirá determinar la carga de una proteína o un péptido en función del pH (ver pH en propiedades de los aminoácidos). A un pH determinado la proteína no tendrá carga neta, ese es el punto isoeléctrico (pI).

Las proteínas aniónicas tienen un pI <7.0 mientras que las catiónicas tendrán un punto isoeléctrico >7.0. (**Figura 3-13**). Esto se traduce en que las proteínas aniónicas tendrán carga negativa incluso a pHs menores que pH 7.0 mientras que las catiónicas tendrán cargas positivas incluso a pHs ma-



También se muestra una micrografía de bacterias E. coli que han expresado proteína recombinante y se ha depositado en forma insoluble en cuerpos de inclusión (fotografía, gentileza del Dr. Rubén Iacono, IDEHU, UBA) y un esquema de la reacción de desplegamiento y solubilización en urea de las proteínas que formaban parte de cuerpos de inclusión. (ver Figura 3-12)

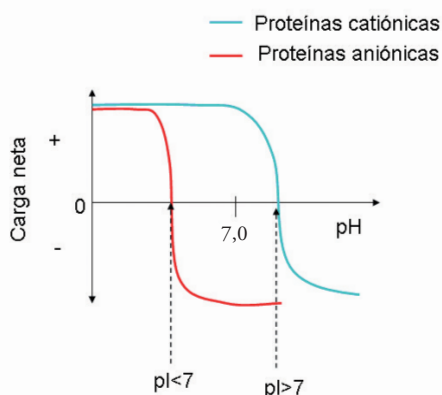


Figura 3-13. La estimación de la carga neta de una proteína en función del pH es importante en el proceso de purificación. Proteínas catiónicas poseen carga positiva a pH superiores a pH 7,0. Las aniónicas poseen carga negativa a pHs menores a pH7,0.

yores a pH 7,0. Esta propiedad podrá ser utilizada para separar proteínas en función de su carga: **las proteínas cargadas positivamente se unirán en forma reversible** a grupos cargados negativamente presentes en las matrices cromatográficas. A su vez, aquellas proteínas cargadas negativamente no se unirán y, por lo tanto, no serán retenidas por este tipo de matrices.

La incubación de la matriz con sales, por ejemplo cloruro de sodio, permitirá la competencia entre las proteínas cargada positivamente y los cationes sodio (Na^+) que funcionarán como contra ión al equilibrar las cargas negativas presentes en la matriz. Si elevamos suficientemente la concentración de sales podremos despegar la proteína unida y eluirla. Entonces la habremos separado los péptidos aprovechando la unión del péptido catiónico (Figura 3-14). Posteriormente, podemos lavar la columna cromatográfica y reestablecer las condiciones iniciales.

Se puede ver que, mediante un cambio en el pH, podemos producir un cambio en la carga neta de la proteína; también podríamos

Como ejemplo sencillo podemos ver lo que ocurre en el caso de la separación de dos péptidos pequeños

Péptido 1

LSKGGLEFLDANLA

Péptido 2

LKKGLKEFLRANLA

¿Cuál será la carga neta de ambos péptidos a pH 7,0?

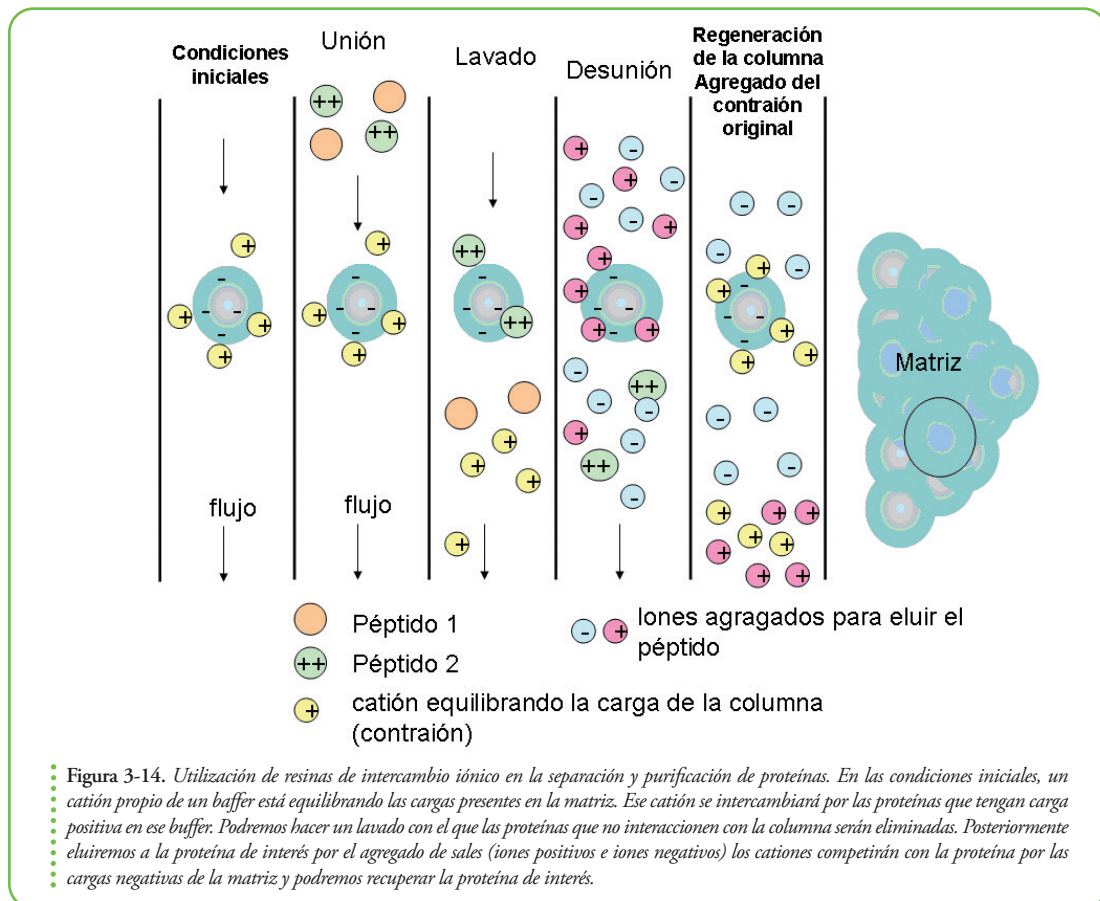
Teniendo en cuenta los valores de pK de cada residuo podemos ver que las lisinas (K) y argininas (R) aportarán una carga neta positiva por residuo mientras que los residuos de ácido aspártico (D) y ácido glutámico (E) aportarán una carga negativa a este pH.

También habrá que tener en cuenta que el grupo amino terminal aporta una carga neta positiva y que el carboxilato terminal aporta una carga neta negativa.

Entonces:

	Péptido 1	Péptido 2
E	1	1
D	1	0
C-terminal	1	1
Q negativas	3	2
K	2	4
R	0	1
N-terminal	1	1
Q positivas	3	6
Q neta	0	+4

Así la carga neta será la suma de las cargas positivas menos la suma de las cargas negativas. A pH 7,0 el péptido 1 tendrá una carga neta $Q=0$ mientras que el péptido 2 tendrá una carga neta $Q=+4$. El péptido 2 quedaría retenido en una matriz de intercambio catiónico, una matriz con grupos cargados negativamente en su superficie, mientras que el péptido 1 no se unirá a esta matriz en este pH.



despegarla y eluirla. Por esto, debemos tener presente que en esta técnica hay dos factores realmente claves.

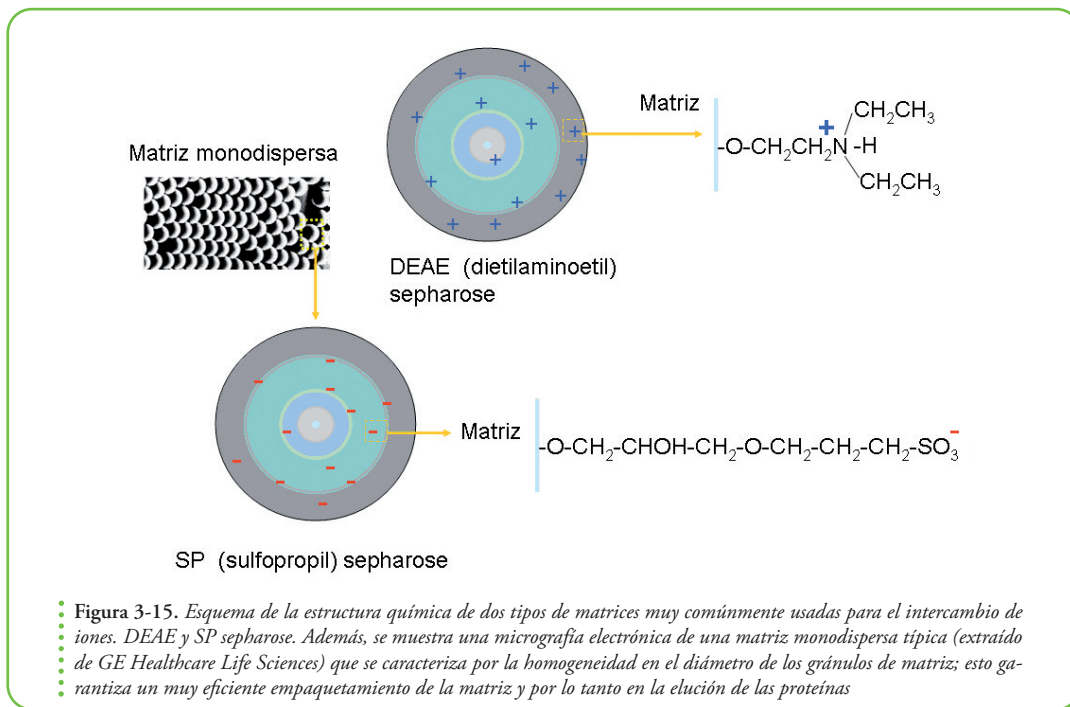
Estos factores son:

- a. la concentración de iones en la solución y,
- b. el pH de la solución.

Existen, en la actualidad, muchas clases de matrices disponibles. En la **Figura 3-15** se muestran dos ejemplos muy usados en el laboratorio, una matriz de intercambio catiónico (sulfopropil) y una matriz de intercambio aniónico (dientilaminoetil).

Si se recurre al intercambio iónico, la proteína purificada eluye, típicamente, con el agregado de sales típicamente en el rango de 100 a 1000 mM. Estas elevadas concentraciones podrían perjudicar estudios posteriores. Existen varias formas de remover sales. Dos de ellas son:

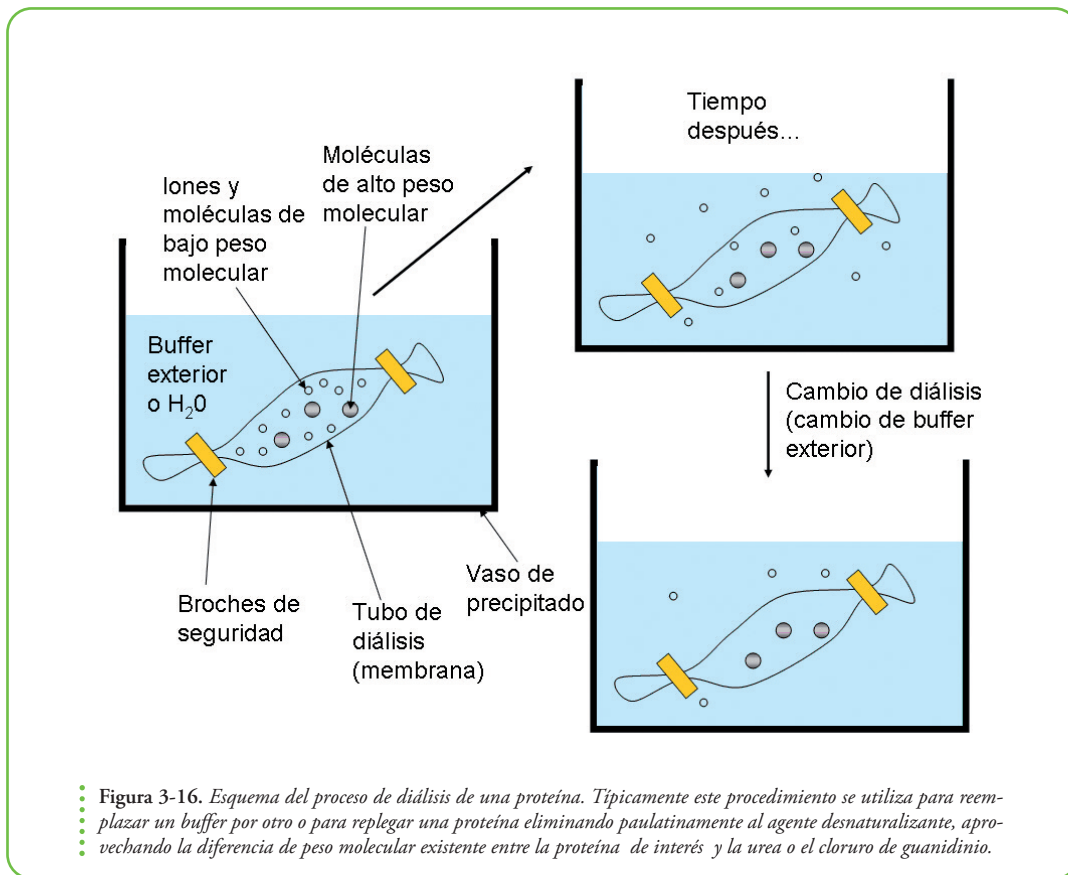
- a. mediante una cromatografía de exclusión molecular (ver Capítulo IV);
- b. la diálisis.



En este último caso, la base del recurso es la utilización de una membrana semi-permeable, la membrana de diálisis que permite la difusión de pequeñas moléculas, como por ejemplo iones sodio o cloruro, pero al mismo tiempo, previene la difusión de moléculas de mayor tamaño. Las moléculas pequeñas difundirán atravesando la membrana hasta que las concentraciones de estas moléculas, dentro y fuera (ver Figura 3-16) del tubo de diálisis sean iguales. En esta situación de equilibrio entrarán y saldrán del tubo de diálisis el mismo número de pequeñas moléculas por unidad de tiempo. Pero si hacemos suficientes cambios del *buffer* exterior podemos disminuir, suficientemente, la concentración de sales.

Es importante destacar que, no solamente podemos eliminar de esta forma iones, también otras moléculas pequeñas. Particularmente, en procesos de replegado nos va a interesar eliminar la urea que usamos para desplegar a la proteína.

Por último, vale la pena mencionar que, la etapa de purificación, es una etapa clave en el estudio de la estructura. Debemos conocer y, esto se logra con experiencia en el laboratorio, las particularidades de cada proteína. La estabilidad conformacional de la proteína de interés dictará los próximos pasos a seguir ya que es clave para el éxito subsecuente determinar cómo conservaremos a las moléculas. Como hemos visto en el capítulo II la desnaturalización es un proceso que ocurre a altas y a bajas temperaturas. No siempre será ventajoso congelar la solución que contiene la proteína. En algunos casos preferiremos mantenerla en presencia de agentes estabilizadores como glicerol al 50 %.



En otros casos podremos conservar a la proteína en condiciones desnaturizantes y replegarla posteriormente para su estudio.

Claro que replegar una proteína no siempre es sencillo. Si bien, como hemos visto el proceso de plegado es un proceso espontáneo, en ciertas condiciones experimentales puede ser muy poco eficiente. Además existen otro tipo de procesos paralelos como el de agregación que compiten y reducen aún más los niveles de proteína correctamente plegada. Por todos estos motivos será crucial conocer aspectos termodinámicos de la proteína de interés y tener a nuestra disposición herramientas que permitan evaluar su estado conformacional.

Estudio de la conformación proteica.

Cuarta Parte

Existe una gran cantidad de técnicas que pueden emplearse en el estudio de las proteínas en el laboratorio. Aquí, nos detendremos en algunas de ellas. Nos importará entender los principios básicos y la teoría que permitirá, luego, aplicar estas herramientas al estudio conformacional de proteínas.

A esta altura, ya hemos comentado varias veces que requerimos, para nuestros experimentos, contar casi siempre, con la proteína de interés relativamente pura y, en general, con una pureza >95%. Por esto, cuando produzcamos una proteína recombinante o cuando aislemos una proteína a partir de una fuente natural (clara de huevo, leche, etc.), vamos a necesitar:

1. Algunas herramientas para poder evaluar el nivel de producción de la proteína recombinante y, posteriormente, evaluar el grado de pureza. Los laboratorios cuentan, típicamente, con sistemas para hacer electroforesis en geles de poliacrilamida (SDS-PAGE por ejemplo) y con un espectrofotómetro de buena calidad para medir la concentración de la proteína pura.

2. Otros equipos y conocer ciertas técnicas que nos permitan purificar la proteína de interés. Típicamente, a estas técnicas las conocemos con el nombre genérico de técnicas cromatográficas. Vamos a discutir los distintos tipos de separaciones por cromatografía: separación por exclusión molecular (separación tamaño) y separación por intercambio iónico (separación por carga).

3. Además vamos a necesitar manejar técnicas que nos permitan reconocer el estado de la proteína de interés. Por ejemplo, durante la producción o durante la purificación podría haberse degradado a través de la acción de proteasas o modificado químicamente, por ejemplo la proteína podría haberse oxidado. Entonces, vamos a necesitar herramientas para evaluar el peso molecular o la masa de la proteína (típicamente espectrometría de masa).

En este punto debemos tener en cuenta que lograr producir en bacterias una proteína recombinante o purificar una proteína de una fuente natural **no garantiza** que su estado conformacional sea el correcto. Por eso, necesitaremos técnicas e instrumentos que permitan evaluar el estado conformacional de las proteínas.

Entre estas técnicas vamos a dedicar unas páginas a discutir algunas espectroscopias muy útiles prestando atención fundamentalmente a cómo interacciona la luz diferencialmente, con cada uno de los estados conformacionales y cómo podemos aprovechar esto para hacer un diagnóstico correcto del estado conformacional de la proteína que estamos estudiando. Absorción UV, absorción de luz circularmente polarizada y fluorescencia, dispersión de luz.

Finalmente, nos introduciremos en una de las dos técnicas complejas que permiten obtener modelos experimentales de proteínas y de alta resolución, cristalografía y difracción de rayos X.

Geles de poliacrilamida

Es una de las técnicas analítica más, comúnmente, usadas en los laboratorios que trabajan con proteínas para la separación y caracterización de mezclas complejas de proteínas y pép-

tidos y para la estimación del peso molecular de este tipo de macromoléculas.

La estimación del peso molecular por geles de poliacrilamida en presencia del detergente SDS (ver más adelante) es, por lo general, muy buena. En la determinación se usan proteínas estándares de peso molecular conocido y comportamiento óptimo (marcadores de peso molecular).

Esencialmente, las partículas cargadas migran en el campo eléctrico hacia el polo de signo opuesto, si tienen carga positiva migrarán hacia el electrodo de polaridad negativa. Las proteínas, como cualquier otra molécula, también migrarán en el campo en función de su carga pero habrá que tener cuidado porque la carga dependerá de la secuencia de la proteína, es decir, de la composición en aminoácidos cargados (esto tiene relevancia, fundamentalmente, si la electroforesis se realiza en ausencia del detergente dodecil sulfato de sodio (SDS), ver más adelante) y, entre otras cosas, de la conformación, del pH (la variación de la carga de la proteína o del péptido con el pH se puede predecir), de la composición de iones del medio electroforético y de la temperatura.

El método más común, para el armado del gel, incluye una reacción de polimerización entre la acrilamida y la N, N'-metilbisacrilamida. Esta reacción es catalizada por persulfato de amonio (APS) y por N, N, N', N'-tetrametiletilendiamina (TEMED) en la que el APS es el iniciador de una reacción catalizada por radicales libres. Podemos pensar que un gel está formado por una trama intrincada con poros pequeños por los que las proteínas deben difundir. Se trata, específicamente, de una matriz de poliacrilamida entrecruzada. La concentración de acrilamida puede variar y, de esto, dependerá el tamaño de los poros y, por consiguiente, el entramado del gel y su poder resolutivo del gel para separar componentes de diversos tamaños.

Independientemente, del efecto de la carga, la movilidad en el gel de poliacrilamida se verá afectada por el tamaño molecular y por la forma de la macromolécula.

Además se suele usar un sistema que permite concentrar las proteínas en un volumen pequeño de tal manera que ingresen todas juntas (en principio) al gel separador. Para esto se prepara un gel que contiene dos fases: una fase concentradora (superior, en color gris en la figura) y una fase separadora (fase inferior). La composición y el pH de ambos geles son diferentes.

Este tipo de experimentos se pueden hacer en ciertas condiciones en las que la carga estará dada por un agente externo (cargado) que tendrá afinidad por las proteínas y se unirá de una manera más o menos homogénea.

El agente más, comúnmente, usado es el SDS, un detergente fuerte que es usado muy frecuentemente para desnaturalizar o desplegar proteínas. SDS desensambla la estructura y disocia los complejos de varias subunidades en encadenas polipeptídicas únicas. El SDS le otorga a todas las proteínas (a casi todas) las mismas propiedades conformacionales e

Puede ser importante tener en mente que el peso molecular de un aminoácido promedio formando parte de una cadena polipeptídica es de unos 113 Da (o 113 g/mol). Así una proteína de 110 aminoácidos tendrá un peso molecular cercano a los 12430 Da o 12.43kDa (12.43 kiloDalton).

El término electroforesis se refiere al movimiento de partículas cargadas (en nuestro caso proteínas) bajo la influencia de un campo eléctrico.

Es muy importante tener en cuenta que la acrilamida es neurotóxica cuando no está polimerizada. Por este motivo SIEMPRE se deben usar guantes y guardapolvo al preparar los geles.

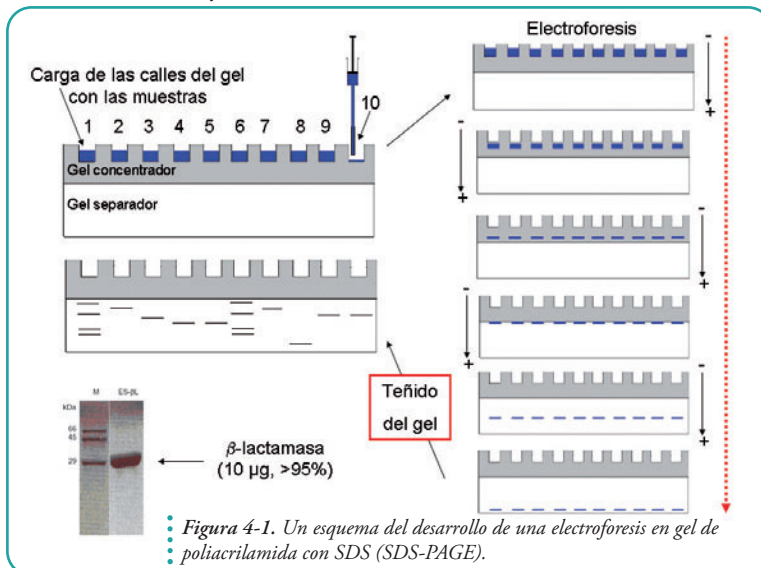
impide que las proteínas interactúen entre sí durante la electroforesis uniéndose a cuanto superficie apolar se presente. El proceso de unión del SDS a las proteínas aun no se conoce a fondo y la naturaleza de la conformación de los complejos SDS: proteína tampoco. Sin embargo los complejos resultantes parecerían tener algunas características comunes: son elongados y poseen flexibilidad conformacional, el largo sería proporcional a la longitud de la cadena polipeptídica.

Además, este detergente enmascara la carga propia de la proteína y le otorga a todas las proteínas carga neta negativa. En general el SDS recubre uniformemente a las proteínas. En general una molécula de SDS se une cada dos aminoácidos en promedio (o 1.4g SDS/1g de proteína). Por esto la carga neta estará, directamente, relacionada con el tamaño.

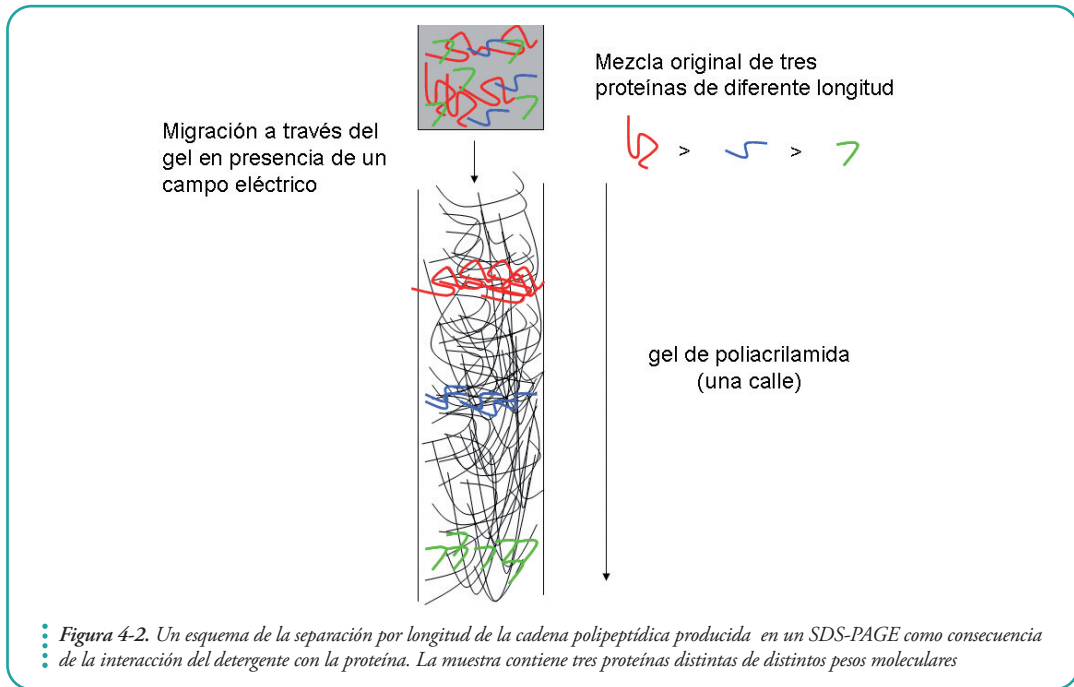
Por lo general, las muestras en un gel con SDS se siembran en presencia de un colorante que migre hacia el polo positivo que no debe interferir en la migración de las proteínas (típicamente se usa azul de bromofenol) para seguir el desarrollo de la corrida (color azul en la **Figura 4-1**). La flecha roja indica el tiempo transcurrido. Este dependerá de la corriente y del voltaje aplicado. Una vez transcurrido el tiempo necesario, se desmontará con cuidado el gel (son muy frágiles) y se procesará para poder visualizar las bandas de proteínas. Con soluciones convencionales de tinte (azul de coomassie por ejemplo) podremos ver sin problemas bandas que contengan como mínimo 0.2 μg . Se verán muy bien bandas que contengan 1 μg de proteína. En nuestra figura hemos sembrado un gel hipotético. En dos de sus calles (calles 1 y 6) sembramos marcadores en el resto varias con proteínas puras. También se muestran dos calles de un gel real en el que se sembró la proteína β -lactamasa (ES β L, 29,5 kDa) luego de su purificación y marcadores de peso molecular conocido (M, con pesos moleculares de 29, 45 y 66 kDa).

Hay que tener presente que si bien, cuanto más grande más cargada estará la proteína, cuanto más grande más le costará a la proteína ingresar por los poros del gel y, en definitiva, más le costará migrar hacia el polo positivo a través de los geles de poliacrilamida.

En la fotografía se muestra el resultado de un SDS-PAGE; se trata de una proteína (ES β L) altamente purificada de peso molecular 29.500 Da. En la calle de la izquierda del gel se sembraron marcadores de peso molecular como referencia. (ver Figura 4-1)



Como consecuencia, los complejos SDS: proteína tienen una movilidad inversamente proporcional al logaritmo del largo de la cadena polipeptídica (Figura 4-2). Por esto mediante la comparación de la movilidad de una proteína desconocida con las movilidades de un conjunto de proteínas de pesos moleculares conocidos podemos inferir el peso molecular de nuestra proteína incógnita dentro del 10 % del valor real.



A pesar de esto, ciertas proteínas se comportan anómalamente. Esto suele estar relacionado, por lo menos en algunos casos, con una distribución de cargas en la superficie particular, estas proteínas suelen ser intrínsecamente muy ácidas o muy básicas, es decir que tienen alta densidad de carga negativa o positiva, respectivamente), y, en otros, con resistencia a la desnaturalización por SDS. En este último caso, se puede agregar a la muestra agente desnaturalizante (urea) para desplegar la proteína de interés. También se debe preparar el gel en presencia de urea. De todas maneras este tipo de geles (SDS-PAGE) nos dará información muy útil relacionada con la longitud de la cadena, pero no en relación a la conformación y a la carga.

Para garantizar la unión del SDS a la proteína, las muestras suelen calentarse a 100 °C en presencia de un exceso de este detergente.

Cromatografía de exclusión molecular

Las técnicas cromatográficas son un conjunto enorme de técnicas fundamentadas en el principio de retención selectiva. Su objetivo es separar los distintos componentes

de una mezcla (en nuestro caso proteínas) permitiendo, con posterioridad y mediante otras experimentos, identificar y determinar las concentraciones de dichos componentes en las fracciones separadas. Las técnicas cromatográficas son muy variadas pero, en todas ellas, hay una fase móvil (fluido, en nuestro caso un líquido) que arrastra a la muestra a través de una fase estacionaria (un sólido o un líquido fijado en un sólido). Las macromoléculas interaccionarán de manera diferente con la fase estacionaria y, así, podrán ser separadas. En particular, la cromatografía de exclusión molecular es un método relativamente sencillo que originalmente se desarrolló para separar proteínas en procesos de purificación.

Sin embargo, este tipo de experimento también permite estudiar, finamente, las propiedades hidrodinámicas de las proteínas: calcular el radio de las mismas asumiendo que son más o menos esféricas. A partir de este dato experimental se puede calcular el peso molecular de la macromolécula. Este tipo de experimento requiere de la elaboración de una curva de calibración que se realizará midiendo el comportamiento de proteínas globulares de peso molecular conocido.

El equipo consta básicamente de:

a. una columna que está cargada con una matriz porosa formada por granos o bolitas de un polímero (tradicionalmente superosa, agarosa o sephadex, entre otros);

b. un sistema de bombas que impulsan la fase móvil (el *buffer*) a través de la columna, si el empaquetamiento no es muy denso puede usarse, simplemente, a la gravedad como fuerza propulsora de la fase móvil;

c. un sistema de inyección de muestra (hay diversos tipos de inyectores, suelen contar con un loop de inyección (que es un tramo de tubería que se carga con la muestra y a través de una válvula queda conectado con la columna y desconectado de la jeringa de inyección);

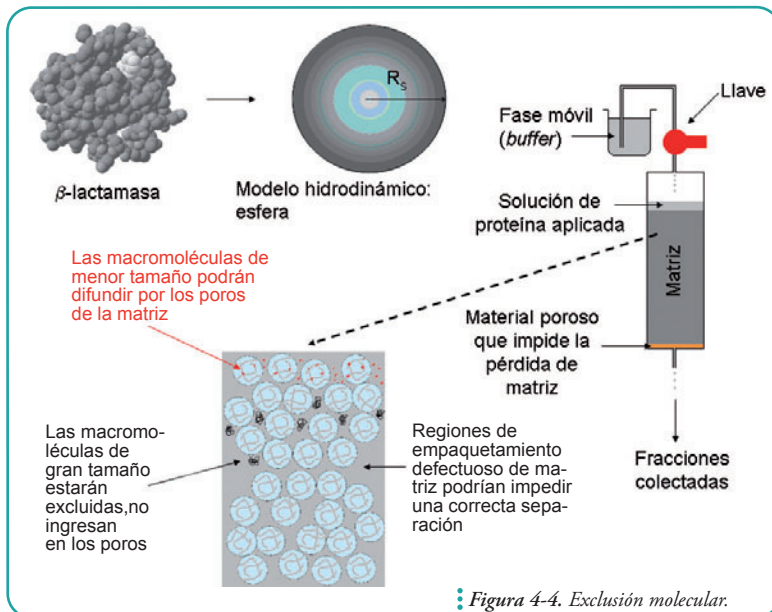
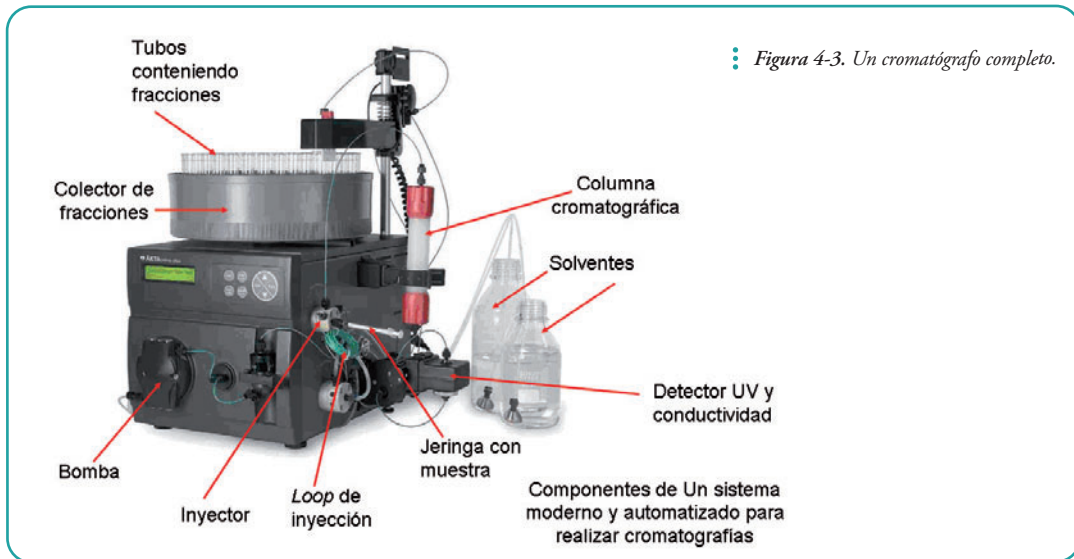
d. un sistema de detección: en nuestro experimento podremos evaluar la elución de las macromoléculas con detectores especiales (debido a que las macromoléculas tienen propiedades espectroscópicas particulares). Por ejemplo podremos medir absorción de luz (por absorbancia típicamente a 280 nm, absorben los residuos aromáticos y a 215 nm absorbe el enlace peptídico, ver más adelante), alternativamente podremos tomar fracciones (muestras) del material eluido y sembrar geles de poliacrilamida y verificar la presencia o ausencia de macromoléculas en cada una de las fracciones (**Figura 4-3**).

El grado de entrecruzamiento y el tipo de polímero permitirá una mejor resolución de la separación de los componentes de la mezcla en los distintos rangos de tamaño molecular.

Parte del espacio dentro de las bolitas de matriz es accesible a ciertas macromoléculas, sólo a las pequeñas, mientras que las de mayor tamaño no podrán ingresar. Así las macromoléculas de mayor tamaño, en promedio, recorrerán caminos más cortos dentro de la columna cromatográfica, en cambio, tanto las más pequeñas como los iones que forman parte de la solución y del regulador de pH (*buffer*) podrán ingresar, recorriendo en promedio caminos más largos por lo que tardarán más en eluir (salir por la parte inferior de la columna, junto con la fase móvil).

El tamaño de una proteína globular es el volumen promedio que ésta molécula ocupa en la solución (Figura 4-4). Podemos pensar, entonces, que existe una esfera que posee un radio determinado cuyas propiedades hidrodinámicas (coeficiente de difusión) serán similares a las de la proteína.

El radio de la esfera es frecuentemente expresado como el Radio de Stokes (R_S). Este parámetro es especialmente sensible a la forma y a las propiedades físicas de la proteína. Si la proteína no es globular el comportamiento será muy diferente al de una esfera.



Arriba, el comportamiento hidrodinámico de las proteínas globulares puede ser modelado por esferas de radio R_S ; la derecha, un esquema típico de una cromatografía en la que se emplea la gravedad para impulsar el solvente (o fase móvil); abajo, un esquema del fundamento de la separación por tamaño molecular. (ver Figura 4-4)

El R_S de dicha proteína puede ser estimado a partir de los datos de la exclusión molecular sólo si la columna cromatográfica utilizada es apropiadamente calibrada con proteínas que tienen aproximadamente la misma forma (también deben ser globulares), las mismas propiedades físicas. Para poder realizar este experimento debemos conocer los R_S y los pesos moleculares de las proteínas utilizadas en la calibración.

Se puede ver que el R_S determinado por técnicas de exclusión molecular se correlaciona extremadamente bien con los valores obtenidos por otros métodos utilizados frecuentemente, como el cálculo a partir de la viscosidad intrínseca de la solución (Uversky, V., *Biochemistry*, 32 (1993), 13.288-13.298).

Para moléculas y proteínas globulares se puede calcular el radio hidrodinámico (R_S , el radio asumiendo que la macromolécula es esférica) como

$$\log R_s = 0,369 \times \log (\text{PM}) - 0,254$$

Donde PM es peso molecular (en Da) de la proteína.

Proteína	Peso molecular (Da)	Volumen de elución (mL)
Tiroglobulina	670.000	8,21
y globulina bovina	150.000	10,64
Ovoalbúmina de gallina	44.000	12,77
Mioglobina equina	17.000	14,06
Vitamina B12*	1.350	19,17

La vitamina B12 se utiliza como marcador por su pequeño tamaño. Este conjunto de marcadores puede ser preparado en el laboratorio.

Así podemos calcular con un mínimo de error los radios de las macromoléculas que usaremos como proteínas y moléculas estándares de tamaño molecular.

En nuestra primer cromatografía podremos inyectar los estándares (inyectamos el conjunto sólo si sabemos que estas proteínas no interaccionan entre sí).

El resultado será un cromatograma como el que se muestra en la **Figura 4-5**. En este cromatograma, cada pico se corresponde con un tipo de proteína o molécula, porque justamente se han elegido proteínas y moléculas que cubran un rango amplio de volúmenes de elución y que no se superpongan.

A partir de nuestro cromatograma podremos medir los volúmenes de elución de cada proteína estándar (**ver Figura 4-5**) para hacer una curva de calibración.

Una vez obtenidos los volúmenes de elución podremos, ya que conocemos los pesos moleculares de estas proteínas, calcular con la ecuación de arriba el radio hidrodinámico para cada una de ellas. Posteriormente si representamos el log (R_S) en función del volumen de elución veremos que hay una dependencia lineal (**ver Figura 4-6**). Esta será nuestra curva de calibración.

Debe quedar claro que si se inyecta una mezcla de proteínas los picos podrían superponerse si es que el comportamiento hidrodinámico fuera similar.

El próximo experimento consistirá en inyectar la proteína de interés, por ejemplo, una proteína cuyo radio hidrodinámico es desconocido. Con este nuevo experimento podremos determinar volumen de elución experimental a partir del cromatograma y posteriormente podremos calcular el radio hidrodinámico usando la regresión lineal obtenida en la curva de calibración (Figura 4-6) experimental.

$$\log (R_s) = 2,6236 - 0,09096 \times V_e$$

Donde V_e es el volumen de elución en mL. Así el radio hidrodinámico experimental para esta proteína calculado a través de cromatografía de exclusión molecular es $R_s = 24,9 \text{ \AA}$.

Es importante ver que, en este tipo de experimentos, debe controlarse muy bien el flujo de la cromatografía. El flujo es el volumen que pasa a través de la columna por unidad de tiempo (mL/min). El flujo permitido dependerá del tipo de polímero y del grado de compactación. Ambos experimentos (curva de calibración y cromatografía de muestra incógnita) deben realizarse en condiciones controladas.

Si conociéramos el valor del peso molecular de la proteína incógnita (por ejemplo 29500 Da) podríamos calcular el **radio hidrodinámico teórico** asumiendo que su forma es globular (que la proteína incógnita es, relativamente, esférica) usando la ecuación de arriba:

$$\log R_s = 0,369 \times \log (29.500) - 0,254$$

$$\log R_s = 0,369 \times 4,469 - 0,254$$

$$\log R_s = 1,395$$

$$R_s = 24,9 \text{ \AA}$$

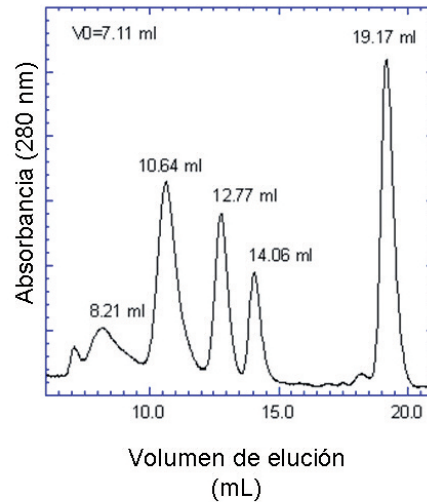


Figura 4-5. Perfil cromatográfico o cromatograma correspondiente a la elución de los marcadores de tamaño molecular; los marcadores fueron sembrados simultáneamente en una mezcla.

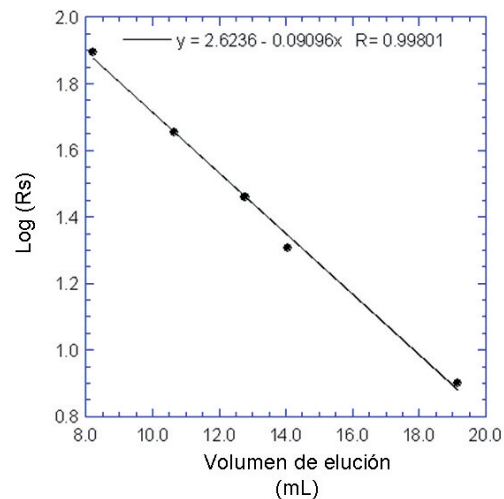


Figura 4-6. Curva de calibración.

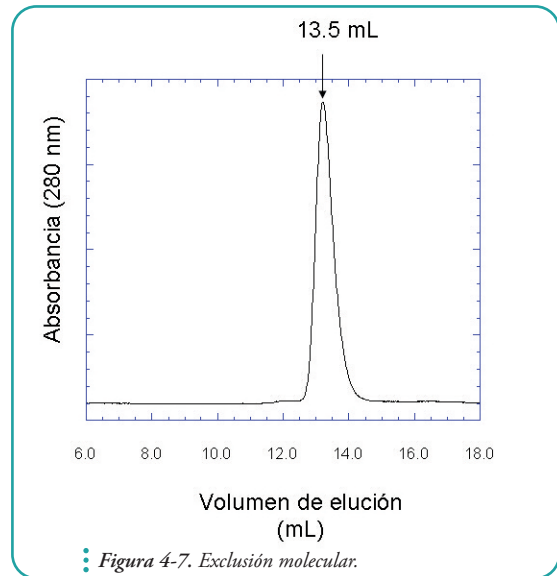
Correlación entre el logaritmo del radio hidrodinámico ($\log R_s$) y el volumen de elución de las proteínas globulares. (ver Figura 4-6.)

Una pregunta pertinente es

¿Por qué hacemos el experimento con la proteína incógnita si podemos calcular, directamente, el radio a través de la correlación entre peso molecular y $\log(R_g)$?

La respuesta es simple. *A priori*, no sabemos si la proteína es globular y tampoco conocemos la estructura cuaternaria de esta proteína. Dicho de otra forma, el resultado de nuestro experimento es, extremadamente, informativo: la proteína incógnita es globular (una esfera es un buen modelo de la proteína) y por otro lado en solución es monomérica.

De la comparación de ambos valores: el experimental y el teórico (en este caso ambos coinciden perfectamente) podríamos calcular el estado de agregación de la proteína: si se trata de un monómero o forma dímeros o tetrámeros por ejemplo.



Cromatograma correspondiente a la β -lactamasa (una proteína globular de 29,500 Da). (ver Figura 4-7.)

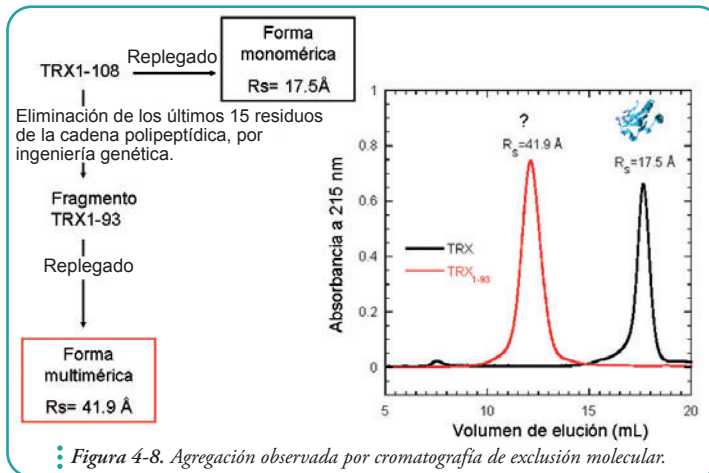
¿Podríamos calcular el R_g en forma teórica de los distintos arreglos cuaternarios posibles?

Subunidades	Arreglo	Peso molecular (Da)	Radio hidrodinámico (Å)
1	Monómero	29,500	24,9
2	Dímero	59,000	¿?
3	Trímero	88,500	¿?
4	Tetrámero	118,000	¿?

La utilidad de este procedimiento es corriente en el laboratorio. Por ejemplo, cuando a una variante de tiorredoxina (que tiene una longitud de 108 residuos) que tiene dos mutaciones (un residuo de ácido glutámico y una glutamina se reemplazaron por histidina) se le eliminó la hélice C- terminal (un elemento de estructura secundaria completo) por técnicas de DNA recombinante se obtuvo un fragmento de menor longitud menor (93 aminoácidos, fragmento TRX 1-93) que después de purificado en condiciones desnaturalizantes se replegó a una forma que, curiosamente, ya no es monomérica, sino, aparentemente, octamérica (Figura 4-8). La agregación es un mecanismo muy común que permite ocultar superficie hidrofóbica que de otra manera se expondría al solvente.

Las cromatografías de exclusión molecular se pueden realizar en condiciones `nativas`, es decir, en solventes en los que va a estar estabilizado mayormente el estado nativo. Pero, también, se pueden llevar a delante en condiciones en las que se establezcan proporciones de

ambos estados (nativo y desplegado) concentraciones intermedias de agentes caotrópicos, urea y GuHCl y, por último, concentraciones elevadas de agentes caotrópicos que estabilizan, únicamente, al estado desplegado, por ejemplo de GuHCl 5-6 M.



El caso de variantes de tiorredoxina. En trazo negro se muestra el cromatograma correspondiente a la proteína salvaje y en trazo rojo el correspondiente a una variante mutante de menor peso molecular pero que tiene alta tendencia a la agregación y forma especies multiméricas. (ver Figura 4-8)

Cabe mencionar que así como existe una correlación entre el logaritmo del peso molecular y el logaritmo del radio hidrodinámico para proteínas en estado nativo también podemos establecer correlaciones para los estados desplegados:

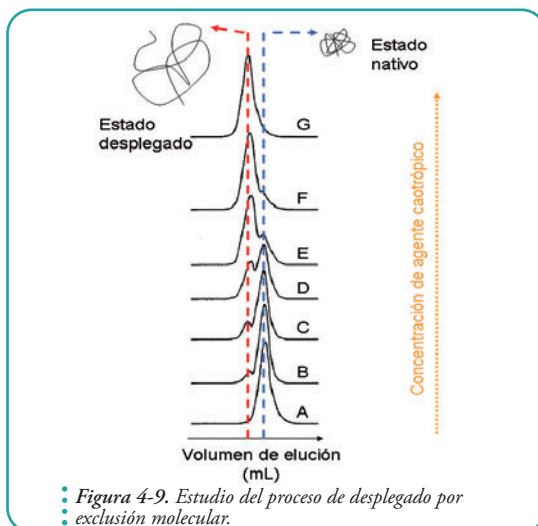
Para proteínas desplegadas en GuHCl vale:

$$\log (R_y) = -0,543 + 0,502 \times \log PM$$

Y para proteínas desplegadas en urea vale

$$\log (R_y) = -0,657 + 0,524 \times \log PM$$

Donde aquí también PM es el peso molecular (en Da) de la proteína. Estas ecuaciones también fueron obtenidas por V. Uversky, 1993.



En el cromatograma (A) el 100 % de las moléculas se encuentran, correctamente, plegadas en el estado nativo. En el cromatograma (G), que se ha realizado a una elevada concentración de GuHCl, (6,2 M) se encuentra estabilizado, únicamente, el estado desplegado que, por ser mucho menos compacto, eluye en un volumen menor. En concentraciones, intermedias (4,1 M) ambas conformaciones coexisten formando dos picos (cromatograma D). La Figura está basada en un experimento realizado con la lisozima de clara de huevo, una proteína sumamente estable y utilizando una columna Superosa 12 (Uversky V. N. Biochemistry 1993,32, 13.288-1 3.298). Para poder hacer estos experimentos es importante demostrar, previamente, que las propiedades físicas de la columna no se alteran en estas condiciones. (ver Figura 4-9)

Figura 4-9. Estudio del proceso de desplegamiento por exclusión molecular.

La existencia de intermediarios o estados, parcialmente, plegados estabilizados en concentraciones intermedias de agentes caotrópicos, también, puede ser evaluada por exclusión molecular. El seguimiento de cambios en el radio hidrodinámico (o el volumen de elución) será nuestra prueba experimental. Así se ha probado la existencia de estados compactos (en los que no cambia el radio hidrodinámico pero, sí cambian otras propiedades del estado nativo, por ejemplo el contenido de estructura terciaria).

Esto es lo que ocurre en el ejemplo que se muestra en la **Figura 4-10**. Se trata de la desnaturalización de la proteína intestinal de unión a ácidos grasos (IFABP). Esta proteína pierde, aproximadamente, el 100% de su estructura terciaria cuando es incubada en GuHCl 2,0 M, sin embargo, en ese rango de concentraciones de desnaturalizante sufre una expansión incompleta: Desde el estado nativo (20 Å de radio) hasta un intermediario (28 Å). Si se incubaba la proteína con mayores concentraciones de desnaturalizante se observará un aumento progresivo del radio hasta alcanzar los 36 Å en el estado desplegado. El estado desplegado teórico se puede modelar con una esfera de radio $R_S = 35,5 \text{ \AA}$.

La desnaturalización se produce en dos etapas: $N \leftrightarrow I$ con una pendiente elevada que se relaciona con un proceso cooperativo de exposición de superficie apolar al solvente y una segunda etapa $I \leftrightarrow U$, un proceso de desplegamiento mucho menos cooperativo.

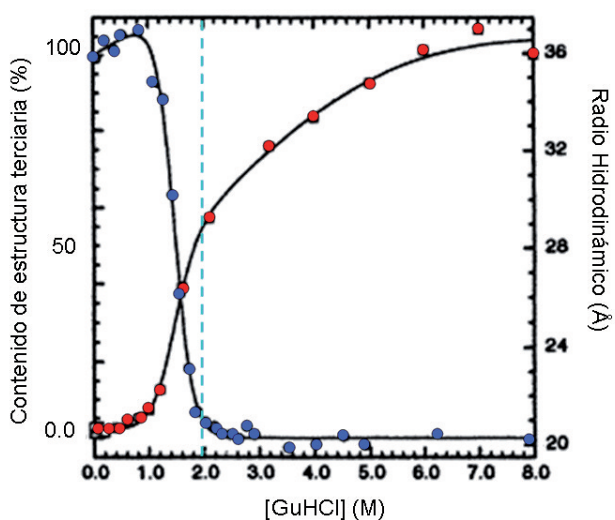


Figura 4-10. Desplegado de la proteína IFABP.

En rojo se muestra la variación en el radio hidrodinámico en función de la concentración de agente desnaturalizante. En azul la variación en el contenido de estructura terciaria. (ver Figura 4-10)

En resumen, hoy en día podemos utilizar este tipo de sistemas para llevar adelante diversos experimentos entre ellos:

- Fraccionamiento por tamaño.
- Separación de monómeros de mezclas con dímeros y agregados de alto peso molecular.
- Estimación del peso molecular
- Estimación del radio hidrodinámico

- Determinación de constantes de equilibrio (monómero \leftrightarrow dímero)
- Desalado, cambios de buffer
- Replegado en columna

Existen diversos tipos de matrices, cada una de ellas con propiedades que las hacen, sumamente, útiles en las distintas aplicaciones.

Otro medio muy, comúnmente, usado en el laboratorio es el Sephadex. Está formado por

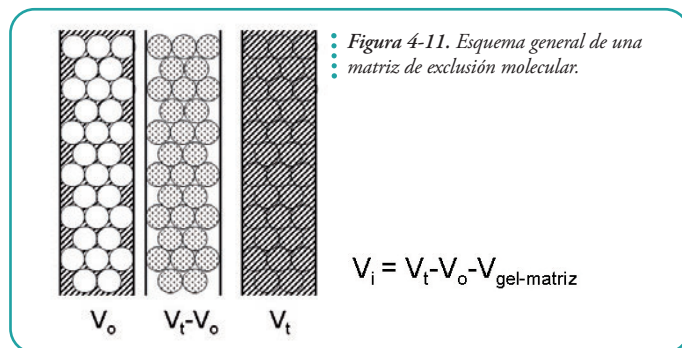
Dextrano (otro polímero) entrecruzado con epíclorhidrina. Es estable en el rango de pH 3-12. También como en el caso de la Superosa se han observado ciertas interacciones hidrofóbicas y el rango de estabilidad térmica es +4-40 °C.

Algunos de los Sephadex son:

Rango de Separación PM (Da)	
Sephadex G 25 Superfino	1.000 - 5.000
Sephadex G 50 Superfino	1.500 - 30.000
Sephadex G 100 Superfino	4.000 - 100.000

Así, por ejemplo, la Superosa consta de partículas de agarosa porosa entrecruzada de $13 \pm 2 \mu\text{M}$. Es estable en el rango de pH 3-12. La Superosa 12 tiene un rango de separación de 1000-300000 Da, mientras que la Superosa 6 tiene un rango de separación que va desde 5000 a 5000000 Da, permitiendo esta última estudiar agregados de alto peso molecular. Es importante tener en cuenta que no soportan temperaturas mayores a 40°C ni menores a 4°C. Algunos investigadores han observado que este tipo de matriz interacciona con regiones hidrofóbicas de algunas proteínas retrasando su elución en la cromatografía.

Por ejemplo, si usamos una columna en la que se ha empaquetado Sephadex G25 y sembramos una proteína de 10.000, una de 30.000 y un compuesto de 250 Da el compuesto de 250 Da eluirá en un volumen mayor porque podrá ingresar en los poros de la matriz mientras que, ambas proteínas, serán excluidas (no ingresarán en los poros) y eluirán en el V_o (ver Figura 4-11) juntas (no se separarán)



Mencionaremos también al Superdex. Es un polímero de azarosa y dextrano. Es una matriz hidrofílica y es estable en el rango de pH 3-12 y de temperatura entre +4-40 °C.

La Superdex S-200 HR tiene un rango de separación que va desde 10000 a 600000 Da. Posee algunas cargas negativas,

por esto, para apantallarlas debe utilizarse alta concentración de sales en la cromatografía. Si se utiliza baja concentración de sales se podrá observar el retraso de proteínas básicas (eluirán en volúmenes de elución mayores a los esperados) y la exclusión de proteínas ácidas (eluirán en volúmenes de elución menores a los esperados para sus pesos moleculares).

Interacción de las proteínas con la luz:

Espectroscopía UV aplicada al análisis conformacional de proteínas

La interacción de la luz con la materia causa transiciones entre distintos estados de energía de las moléculas.

La espectroscopia de absorción UV (ultravioleta) involucra la absorción de luz UV por parte de una molécula promoviendo el paso de un electrón (por esto decimos que son tran-

siciones electrónicas) desde un orbital molecular fundamental (los orbitales moleculares son el producto de la combinación de los orbitales atómicos) a un orbital excitado. La radiación aporta la energía suficiente para que se den las transiciones electrónicas entre orbitales. La separación energética entre estos orbitales moleculares corresponde a las frecuencias y longitudes de onda del UV.

Podemos recordar que la longitud de onda (λ) es inversamente proporcional a la frecuencia (f)

$$\lambda = \frac{c}{f}$$

Donde la constante c vale 299.792.458 m/s (la velocidad de la luz) para las ondas electromagnéticas que viajan en el vacío.

Cuando las ondas electromagnéticas penetran en un medio, como puede ser una solución acuosa (en la mayor parte de nuestros casos), la longitud de onda se reduce de forma proporcional al índice de refracción (η) de dicho medio, la frecuencia disminuye así la constante de la velocidad de la luz queda invariante.

$$\lambda_1 = \frac{\lambda_0}{\eta}$$

Donde λ_0 es la longitud de onda en el vacío y λ_1 es la longitud de onda en el nuevo medio (en la solución acuosa, por ejemplo). La longitud de onda de las radiaciones electromagnéticas, se expresa, por lo general, en función de la longitud de onda de estas radiaciones en el vacío y, en nuestro caso, las unidades de medida más comúnmente usadas serán los nanómetros ($1\text{nm} = 10^{-9}\text{m}$).

También es importante recordar que la frecuencia (y la longitud de onda) estará asociada con la energía de la onda electromagnética.

Los orbitales moleculares serán ocupados cada uno por dos electrones. El último orbital ocupado recibe el nombre genérico de orbital HOMO y el primer orbital desocupado, que es una región del espacio que tiene menor probabilidad que HOMO de ser ocupada por electrones, recibe el nombre genérico de orbital LUMO; LUMO es el orbital molecular desocupado de menor energía. En la **Figura 4-12** los electrones están representados por pequeñas flechas de color verde.

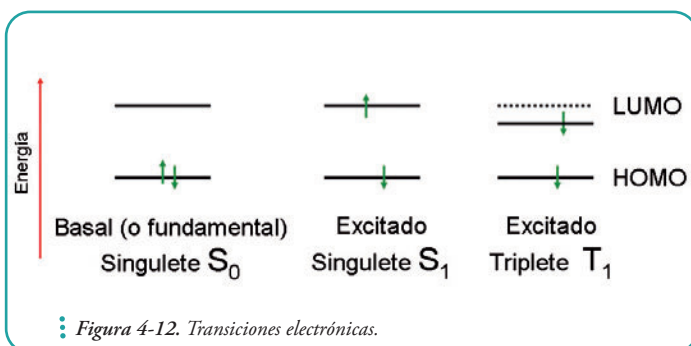


Figura 4-12. Transiciones electrónicas.

La absorción de energía promueve la ocupación del orbital LUMO con un electrón. Debemos notar que el sentido de las flechas verdes (arriba y abajo) no es arbitrario sino que por ciertas cuestiones de la física cuántica dos electrones no pueden ser idénticos (no pueden poseer los mismos números cuánticos). El sentido de las flechas está asociado al spin del electrón en cuestión. (ver Figura 4-12).

El proceso de absorción de energía (dada la interacción con una onda electromagnética de determinada frecuencia) por parte de un electrón y su promoción a un orbital LUMO (estado excitado) es un proceso muy rápido y está en el orden de 10^{-15} segundos (0.000000000000001 segundos).

Lo realmente interesante es que podemos medir este proceso (la absorción de energía por parte de electrones) mediante técnicas y equipamientos relativamente sencillos y comunes. Además, nos interesará, particularmente, porque la absorción estará relacionada con la cantidad de materia que absorba energía. Dicho de otro modo será entre otras cosas una excelente herramienta para determinar la concentración de proteínas en una solución.

Para esto debemos usar una cubeta transparente a la radiación, así podremos medir la intensidad del haz de luz una vez que ha atravesado la cubeta que contiene sólo *buffer* y, posteriormente, podremos medir la intensidad del haz de luz que ha atravesado la cubeta pero, ahora, presencia de la proteína: disolvemos la proteína en la misma solución *buffer*.

Se puede ver que

$$\frac{I_f}{I_0} = 10^{-\epsilon \cdot l \cdot C}$$

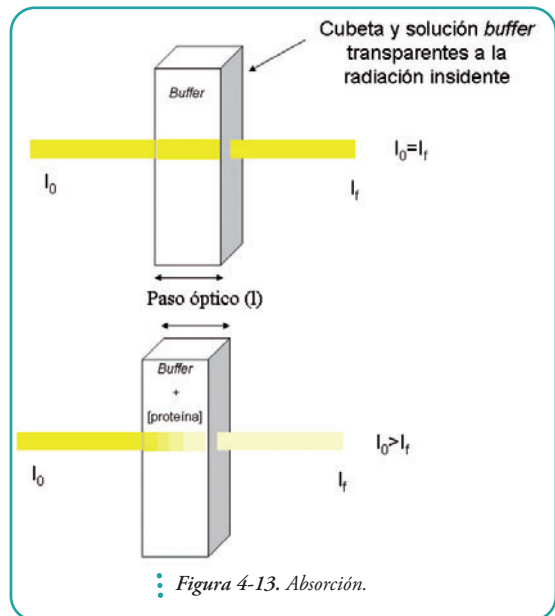
Donde I_f es la intensidad a la salida de la cubeta e I_0 es la intensidad en ausencia de absorción. El paso óptico de la cubeta es l (el tramo, por ejemplo en centímetros que atravesará la luz) y C es la concentración de la sustancia que absorbe. ϵ se conoce con el nombre de coeficiente de extinción y variará dependiendo de la longitud de onda y del tipo de sustancia y de su composición. En el caso de las proteínas variará dependiendo, por lo general, de la secuencia de aminoácidos de cada proteína, y de la longitud de onda a la que se realice el experimento.

En el laboratorio suele emplearse la medida de absorbancia porque la absorbancia es proporcional al paso óptico y a la concentración de la sustancia disuelta.

$$A = -\log \frac{I_f}{I_0}$$

Y por lo tanto, teniendo en cuenta la ecuación de más arriba,

Así, conociendo el coeficiente de extinción, podremos calcular la concentración si es que podemos



Arriba, una muestra que no absorbe luz, en efecto, la intensidad I_0 antes de atravesar la muestra es idéntica a la intensidad luego de atravesarla; en cambio, abajo la intensidad luego de atravesar la muestra es menor debido al proceso de absorción de luz (I_f). (ver Figura 4-13)

medir, experimentalmente, la absorbancia (Ley de Lambert-Beer). Casi todos los laboratorios en los que se trabaja con proteínas cuentan con un espectrofotómetro, este es el equipo que permite hacer estas mediciones.

Para poder determinar la concentración tendremos que contar con un coeficiente de extinción correctamente determinado.

Muchas proteínas contienen además de los residuos de amino ácidos otros grupos que pueden encontrarse unidos a ellas tanto de forma covalente como no covalentemente. Y en muchos casos, puede que estos grupos sean cromóforos y contribuyan al espectro de absorción de la proteína en la región del espectro que se conoce como UV-cercano (240-340 nm) o en el visible (400-750 nm).

La determinación correcta de la concentración proteica será indispensable para poder validar todos los estudios conformacionales que se le realicen a cada proteína porque las señales que determinaremos asociadas al contenido de estructura secundaria y terciaria, por lo general, también dependerán de la concentración proteica.

Algunos ejemplos de proteínas con grupos de este tipo son la hemoglobina que tiene un grupo hemo por subunidad, las proteínas con centros de hierro-azufre, y las flavoproteínas (que cuentan con grupos flavinas).

Una cuestión importante, que seguramente han notado, es que el equipo que permite hacer estas mediciones (el espectrofotómetro) debe ser capaz de recorrer un rango amplio de longitudes de onda (debe hacer un *barrido*), de tal forma que la muestra sea iluminada con cada una de las longitudes de onda. Para cada longitud de onda, se medirá la absorbancia.

En el caso de que la proteína no cuente con este tipo de grupo y tampoco tenga centros metálicos especiales (como los de unión a Zn, Ni, Mg o Cu, por ejemplo) podremos afirmar que la contribución a la absorción de luz estará dada por la cadena polipeptídica en si misma. Como veremos más adelante la estructura que esta adquiera en solución tendrá importantes efectos.

Un punto importante es identificar a los grupos que absorben luz en el rango de longitudes de onda que estudiamos:

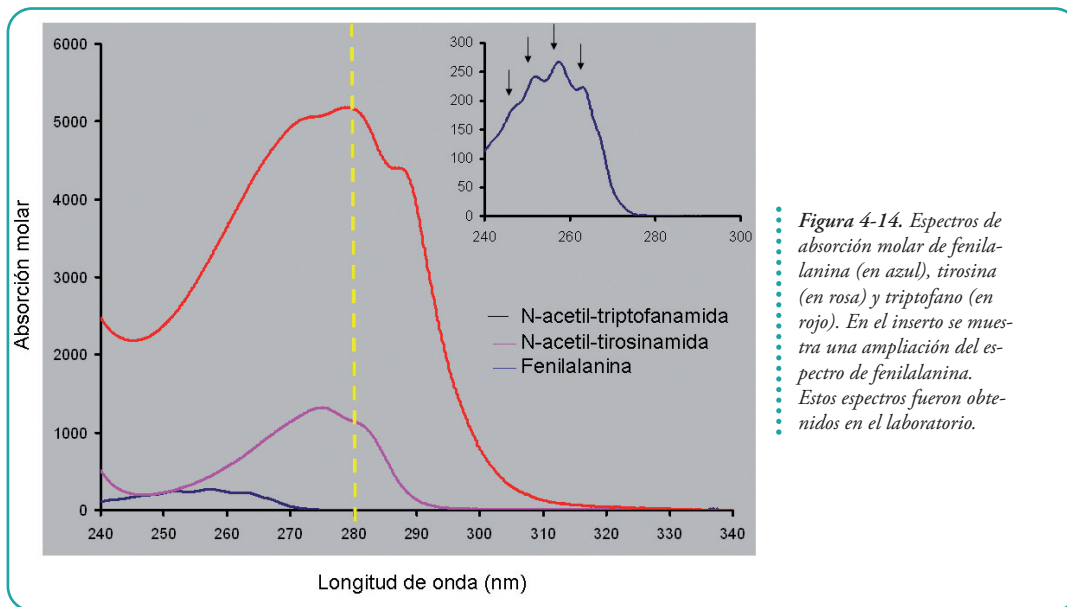
El ultravioleta lejano: El enlace peptídico.

En la región del ultravioleta lejano (entre 170 y 220 nm) absorbe el enlace peptídico. Existen dos tipos de transiciones electrónicas: electrones que se encuentran en orbitales llamados n se excitan y se promueven a orbitales que se conocen con el nombre de orbitales π^* (estas transiciones se observan a 220 nm). Por otro lado, electrones que se encuentran en orbitales π (son una combinación de orbitales atómicos p perpendiculares al enlace) también pueden promoverse a orbitales π^* , estas últimas son transiciones que se observan a 190 nm y producen bandas de absorción en esta región que serán predominantes.

Estas transiciones se basan en la teoría de los orbitales moleculares; los orbitales σ , π y n corresponden al estado basal o fundamental y los orbitales σ^* , π^* y n^* por el contrario son, como ya se dijo, de mayor energía y reciben el nombre de antienlazantes. Las transiciones $\sigma \rightarrow \sigma^*$ requieren mucha energía, y no serán observadas por métodos ópticos.

El ultravioleta cercano: Los residuos aromáticos.

En la región del ultravioleta cercano (entre 240 y 340 nm), en cambio, aportan al espectro de absorción UV los residuos aromáticos de triptofano, tirosina y fenilalanina. Por esto el espectro UV en esta región dependerá, básicamente, de la composición de aminoácidos. También tienen un pequeño aporte los puentes disulfuro (S-S).



En la **Figura 4-14** se muestra el aspecto de los espectros de *N-acetil-triptofanamida* (NATA), *N-acetil-tirosinamida* (NAYA) y fenilalanina (Phe). Estos compuestos, NATA y NAYA se utilizan como compuestos modelo de residuos de triptofano y tirosina inmersos en un entorno similar al enlace peptídico aportado por la cadena polipeptídica. Son los espectros de soluciones 1.0 M de cada compuesto aislado. Estos espectros se calculan en base a soluciones diluidas ($0.5 < \text{absorbancia} < 1$, máximo de longitud de onda), teniendo en cuenta los valores de los coeficientes de extinción molar (**Tabla 4-1**)

Coeficientes de extinción molar ($\text{M}^{-1} \times \text{cm}^{-1}$)	
<i>N-acetil-triptofanamida</i> (NATA)	5.390,00 (280 nm)
<i>N-acetil-tirosinamida</i> (NAYA)	1.390,00 (275 nm)
<i>Fenilalanina</i> (Phe)	284,57 (257,4 nm)
<i>Puente disulfuro</i> (S-S)	120,00 (280 nm)

Tabla 4-1. Coeficientes de extinción molar de grupos funcionales que absorben significativamente en la región del UV cercano (240-340 nm)

Así, la absorbancia a cada longitud de onda de estas soluciones será la absorción molar

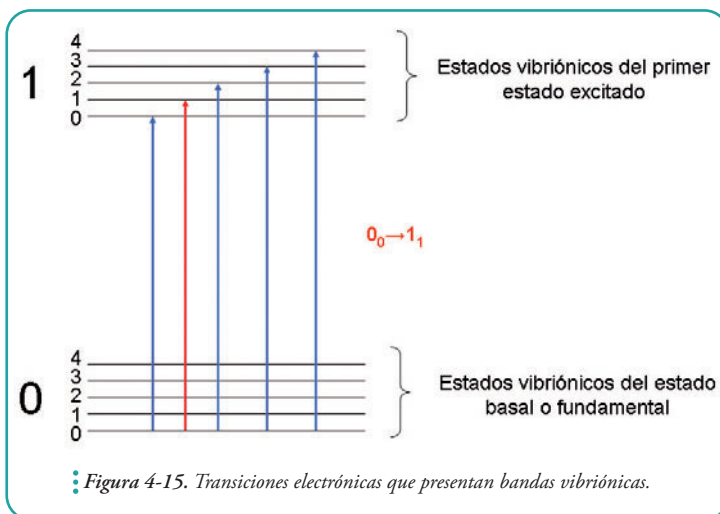
ya que se utiliza una cubeta de paso óptico 1 cm y las unidades de los coeficientes de absorción molar usados son en este caso $M^{-1}cm^{-1}$. El resultado es el valor de absorbancia (adimensional) ϵ .

$$A = \epsilon \times C \times l$$

$$A = \epsilon (cm^{-1} M^{-1}) \times 1 cm \times 1 M$$

$$A = \epsilon$$

Lo primero que resalta de estos espectros es que el espectro de NATA (nuestro modelo de triptofano) prevalece en todo el rango de longitudes de onda sobre los espectros de NAYA (modelo de tirosina) y al de fenilalanina. Particularmente, puede verse que en el rango de 295 a 300 nm Trp es el único residuo aromático que absorbe. Otro punto importante es que a 280 nm las fenilalaninas no absorben luz. En el laboratorio es muy común medir absorbancia a 280 nm como indicador de concentración; sabemos ahora que los residuos que aportan a esta longitud de onda son las tirosinas y los triptofanos. Por otro lado, en el caso de la fenilalanina pueden verse con mucho detalle varias bandas de absorción (el inserto en la **Figura 4-14**) señaladas con flechas. Estas bandas se conocen como bandas vibrónicas y surgen del efecto de las vibraciones en los enlaces del anillo aromático sobre la energía de los orbitales moleculares. Cada una de estas bandas corresponde a un nivel vibrónico distinto (ver **Figura 4-15**) y tiene asociada una frecuencia (una longitud de onda) y una energía. Aquellas bandas en las que la longitud de onda es mayor (menor frecuencia) tendrán asociada una energía menor.



Por ejemplo la transición $0_0 \rightarrow 1_0$ requerirá menor energía que la transición $0_0 \rightarrow 1_1$; aquí con los subíndices se denota los estados vibrónicos. (ver Figura 4-15)

Una pregunta sensata es

¿Qué pasa si la proteína de interés no posee en su secuencia ni tirosinas ni triptofanos?

Tendremos que ingeniar otros experimentos para poder calcular la concentración de proteína y no podremos realizar gran parte de los experimentos que incluyen pruebas sobre este tipo de residuo. Alternativamente, podríamos incluir por ingeniería genética

algún residuo aromático en la secuencia.

Otra pregunta que podemos hacernos es si es posible calcular el espectro de una proteína en base a la composición de aminoácidos de ésta. La respuesta es que sí, se puede.

Se puede modelar el espectro de una proteína desplegada en base a una suma de espectros de los residuos aromáticos y de los puentes disulfuro presentes en presencia de desnaturalizante.

¿Cuál será el espectro de la proteína β -lactamasa de *Bacillus licheniformis* que tiene en su secuencia tres triptofanos, siete fenilalanina, seis tirosinas y no tiene puentes disulfuro?

Para entender esta idea, podemos comenzar por contestar cuál será el coeficiente de extinción a 280 nm para esta proteína en estado desplegado.

Si conocemos la composición de residuos de la proteína y el número de puentes disulfuro (si es que la proteína tiene cisteínas formando puentes disulfuro), entonces podremos calcular el coeficiente de extinción.

$$\epsilon_{280nm} = n_{Trp} \times \epsilon_{Trp280nm} + n_{Tyr} \times \epsilon_{Tyr280nm} + n_{Phe} \times \epsilon_{Phe280nm} + n_{S-S} \times \epsilon_{S-S280nm}$$

	Urea 8 M	GuHCl 6M
Triptofano	5.635	5.685
Tirosina	1.300	1.285
S-S	115	125
Phe	0	0
Longitud de onda: 280 nm		

Si calculamos el coeficiente de extinción molar teniendo en cuenta los coeficientes de extinción de los compuestos modelo en presencia de alta concentración de agente caotrópico (en este ejemplo GuHCl 6M), el coeficiente de extinción de esta proteína será

$$\begin{aligned} \epsilon_{280nm} &= 3 \times 5.685 + 6 \times 1.285 + 7 \times 0 + 0 \times 125 \\ &= 24.565 M^{-1} cm^{-1} \end{aligned}$$

Efectivamente, éste es el coeficiente de extinción de la proteína desnaturalizada en GuHCl 6M.

En el estado nativo las bandas de absorción se suavizan porque cada residuo de la molécula está en un ambiente particular originando pequeñas diferencias y, a su vez, todas las bandas de absorción se superponen.

Algunos de los ambientes son más polares que otros y, algunos, serán definitivamente ambientes apolares, por ejemplo los residuos que estén en el corazón de la proteína. Una de las causas de la existencia de estas

Un punto importante es que todos los residuos en el estado desplegado estarán, por lo menos, en principio, igualmente expuestos al solvente. Esto no ocurre en el estado nativo, en el que cada residuo se encuentra en un ambiente particular. Por este motivo, el espectro del estado nativo será algo diferente al del estado desplegado y, por lo tanto, los coeficientes de extinción también.

diferencias entre los espectros de N y U es el cambio en la polaridad del entorno químico inmediato del residuo que absorbe. A pesar de que, en algunos casos, las diferencias son pequeñas (e incluso hay que hacer procesamientos matemáticos de derivación de los espectros) pueden ser lo suficientemente significativas para ser medidas y estudiadas. Estudiando estas particularidades podremos inferir propiedades del entorno de los aminoácidos en la estructura nativa. También podremos distinguir si una proteína está correctamente plegada o se encuentra desplegada o parcialmente plegada. En la **Figura 4-16** se muestra el espectro (absorción molar) de la proteína β -lactamasa. Puede verse que la región correspondiente a la fenilalaninas presenta ciertas irregularidades que se deben a las bandas vibrónicas. Además, en la región de superposición de Tyr y Trp se observan bandas que son características del estado nativo (en el inserto se amplía esta región del espectro). Con una flecha roja se muestra una de las bandas de absorción propias del triptofano (ver también **Figura 4-14**).

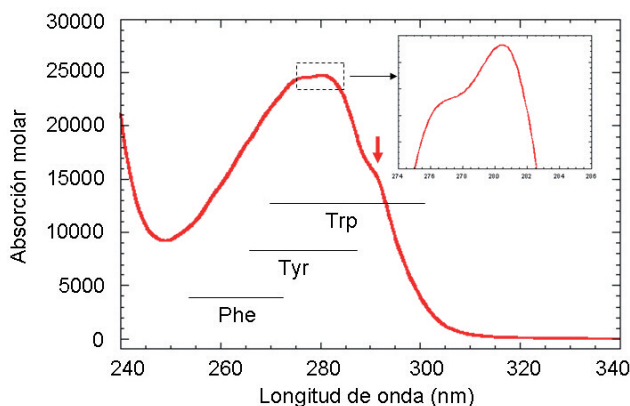


Figura 4-16. Asignación de regiones espectrales correspondientes a los residuos aromáticos.

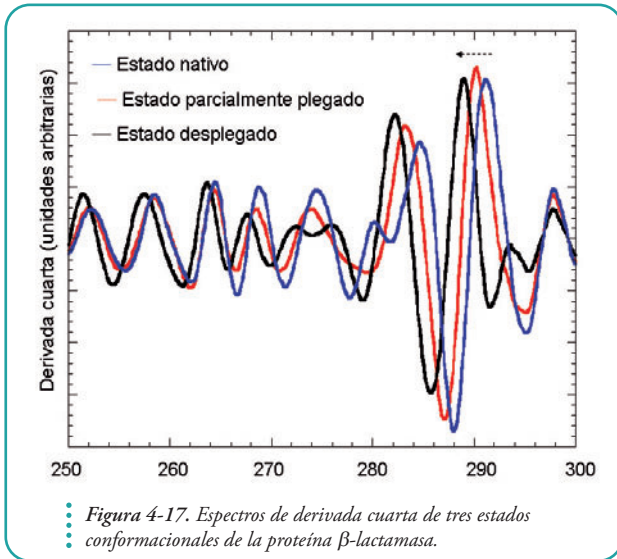
En el inserto se amplía la región del espectro que corresponde a la superposición de Tyr y Trp, aquí se observan bandas que son características del estado nativo. Con una flecha roja se identifica una de las bandas de absorción de triptofano. (ver Figura 4-16)

Las bandas de absorción, en el estado nativo, suelen estar desplazadas con respecto a las bandas de absorción del estado desplegado y, esto, suele ser muy notorio para los triptofanos; no siendo así para las tirosinas y, menos aún, para las fenilalaninas que son mucho menos sensibles al cambio de polaridad en el entorno químico.

Una de las maneras más sencillas para ver estas diferencias espectrales es derivar numéricamente los espectros. Este procedimiento y, en particular la derivación cuarta (que es sencilla de analizar porque sus máximos coinciden con los del espectro original), permite detectar cambios sutiles, corrimientos cercanos al nanometro.

En la **Figura 4-17** se muestra, como ejemplo, los espectros de derivada cuarta de tres estados conformacionales de la proteína β -lactamasa, el estado nativo, un estado parcialmente plegado obtenido mediante ingeniería genética de la proteína (que no altera la composición de residuos aromáticos de la proteína) y el estado desplegado modelado, como la suma de las contribuciones de los residuos de la proteína.

Podemos analizar el corrimiento de bandas de absorción desde un punto de vista algo más teórico. Como dijimos más arriba el espectro de absorción de una cadena polipeptídica plegada no es, exactamente, la suma de los espectros de los grupos que la constituyen.



En azul el estado nativo, en negro el estado desplegado y en rojo un espectro correspondiente a una conformación parcialmente plegada. (ver Figura 4-17)

Podemos observar dos tipos de corrimiento para una banda de absorción: desplazamiento hipsocrómico y desplazamiento batocrómico. Estos desplazamientos que ocurren hacia una longitud de onda menor o mayor, respectivamente, se deben a cambios en la energía del estado fundamental y en la energía del estado

excitado. Estos cambios en la energía de los orbitales moleculares dependen de la polaridad del entorno y de la distribución de carga en estos estados.

En la **Figura 4-18** se muestran dos transiciones electrónicas en las que ambos estados, el excitado y el basal, se estabilizan en el entorno polar pero estabilizará aún más el estado electrónico en el que la separación (distribución) de carga sea máxima. Cuando hay electrones no compartidos (electrones n), pueden formarse puentes de hidrógeno con el solvente polar, entonces, la transición requiere más energía (efecto hipsocrómico), en definitiva observaremos que la longitud de onda se desplaza a valores menores (menor longitud de onda, mayor energía asociada). En cambio en las transiciones π a π^* los estados excitados son más

polares que los fundamentales y, por esto, un aumento en la polaridad del solvente hará que, por la interacción entre dipolos, disminuya más la energía del estado excitado que la del fundamental. Como consecuencia se requerirá menor energía para provocar esta transición en un solvente polar y la transición tendrá asociada una longitud de onda mayor (efecto batocrómico).

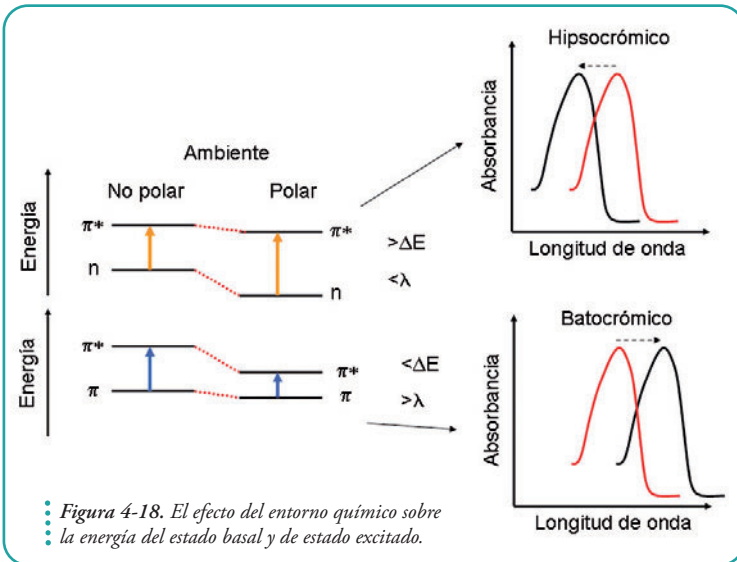


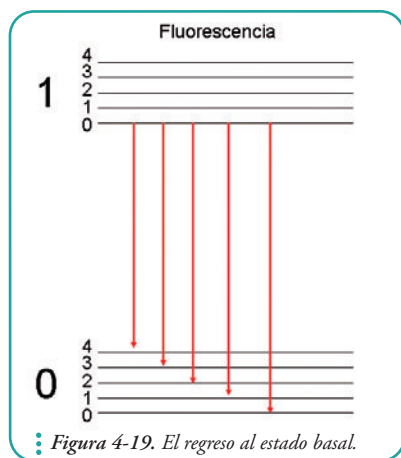
Figura 4-18. El efecto del entorno químico sobre la energía del estado basal y de estado excitado.

Fluorescencia

Hemos visto anteriormente cómo es que podemos excitar electrones y cómo podemos detectar el proceso de absorción de energía. También hemos visto que podemos sacar provecho de este fenómeno y obtener información acerca de una muestra de proteína en solución acuosa. Todavía más interesante, podemos obtener información sobre el estado conformacional de estas macromoléculas.

Una pregunta importante es cómo se desexcita el estado excitado. Y, yendo más a nuestro tema, cómo podemos aprovechar el proceso de desexcitación (la vuelta al estado basal o fundamental de los electrones) para obtener más información sobre el estado en que se encuentran las moléculas en la muestra.

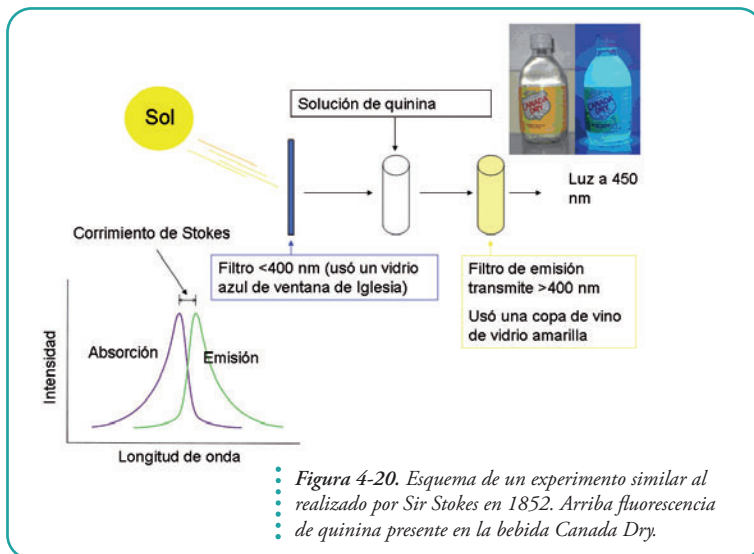
Fluorescencia es uno de los procesos por los que una molécula excitada puede volver a su estado fundamental (de menor energía), en este caso la vuelta se hace con emisión de luz (emisión de un fotón). Este fotón que se emite en el proceso de desexcitación podrá ser detectado. El equipo que usaremos para su detección es el fluorómetro. Como las moléculas, una vez excitadas, se *relajan* (vuelven) rápidamente al estado vibracional de menor energía del primer estado electrónico excitado (por procesos que se conocen con el nombre de conversión interna) **la emisión de fluorescencia ocurre únicamente desde el estado excitado de menor energía** y esto ocurre porque los estados excitados de mayor energía se relajan al estado excitado de menor energía más rápidamente (en el orden del pico segundo, 10^{-12} s) de lo que tardarían en emitir un fotón (en el orden del nano segundo, 10^{-9} s) (ver **Figura 4-19**).



La emisión de fluorescencia ocurre, únicamente, desde el estado excitado (1) de menor energía. Pueden observarse los subiveles vibrónicos correspondientes al estado basal (0). (ver Figura 4-19)

Esto se conoce como regla de Kasha, por el investigador que Michel Kasha y el proceso que determina un corrimiento de la longitud de onda de emisión de luz con respecto a la de absorción (corrimiento de Stokes) había sido descrito con mucha anterioridad (en el año 1852) por Sir G.G. Stokes quien realizó el experimento que se muestra en la **Figura 4-20**.

Sir Stokes muy ingeniosamente ubicó una serie de elementos ópticos de tal forma que pudo comprobar que, si iluminaba una solución sólo con longitudes de onda menores a 400 nm, existía un proceso por el cual la materia absorbía luz y emitía, posteriormente, radiaciones de longitudes de onda mayores a 400 nm (450nm). Para esto Sir Stokes usó, como elementos ópticos, ingeniosos filtros: un cristal azul de la ventana de una Iglesia que no dejaba pasar luz de longitudes de onda mayores a 400 nm y una copa de vino de cristal amarillo que impedía el paso de luz con longitudes de onda menores a 400. En definitiva si la sustancia no hubiera sido fluorescente, Sir. Stokes no hubiera observado

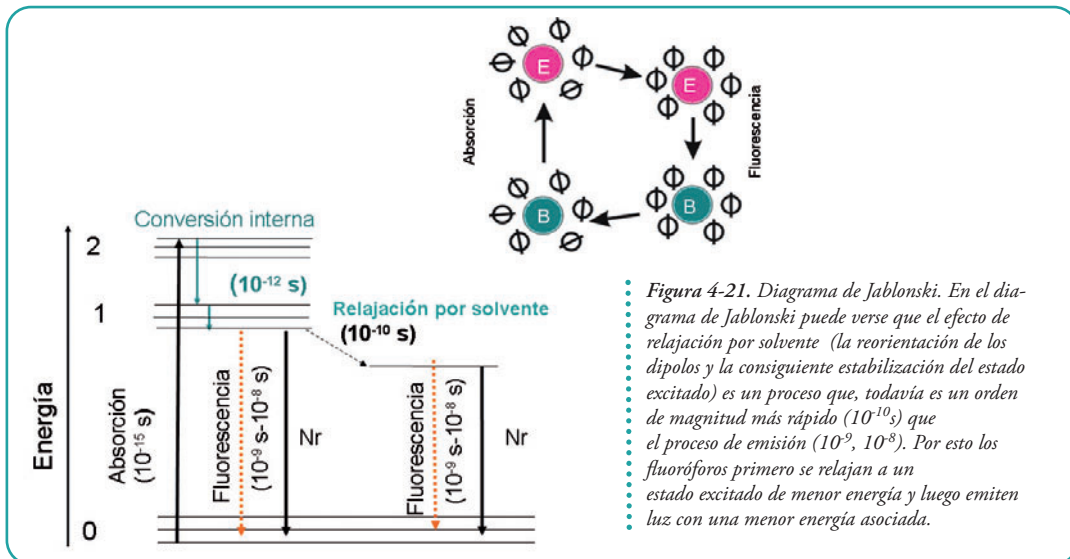


luz a través de la copa amarilla.

Existen otros procesos que modulan la longitud de onda de emisión de fluorescencia. Uno de ellos, que es importante para el estudio de la conformación de proteínas, es el de relajación por solvente. La reorientación de los dipolos de las moléculas de agua (solvente) entorno al estado excitado del fluoróforo (fluoróforo es toda molécula que tiene la capacidad

de fluorecer) puede disminuir la energía del estado excitado y, por lo tanto, el fotón emitido tendrá asociada una longitud de onda mayor (menos energética). En otras palabras, este reordenamiento de dipolos puede estabilizar al estado excitado. Este efecto se conoce como relajación del solvente. Los fluoróforos, que poseen un estado excitado con una mayor separación de cargas con respecto al estado basal, son más sensibles a este efecto, ya que las moléculas de solvente (agua) se reorientarán en su entorno. En la Figura 4-21 se muestra un esquema de este efecto (en el que B indica estado basal y E indica estado excitado) y el diagrama de Jablonski.

Es importante, a esta altura, que quede claro que la emisión es una de las formas de desexcitación. La desexcitación puede ocurrir por otras vías, entre estas otras, por

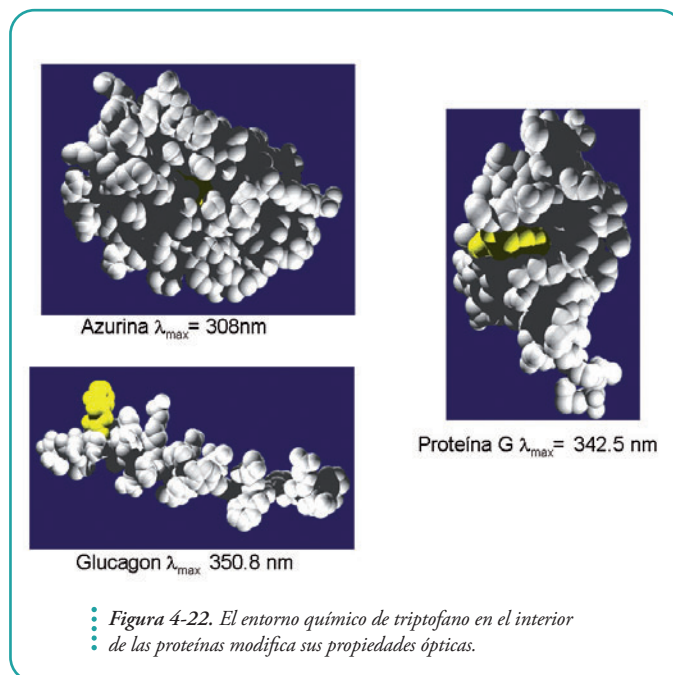


transferencia de esa energía a otras moléculas, colisiones, vibraciones; estas vías son no radiativas (no tienen asociada emisión de luz). Como veremos no todas las moléculas son fluoróforos. Los fluoróforos tienen propiedades químicas especiales.

Tienen sistemas de electrones conjugados.

Para las proteínas los residuos de tirosina y triptófano son fluoróforos útiles en el rango de luz ultra violeta. En particular los residuos de triptófano son excelentes sondas para estudiar la conformación proteica por fluorescencia: el triptófano es muy sensible al entorno (por efecto de relajación de solvente).

Es posible ver cambios en los espectros de emisión en respuesta a:

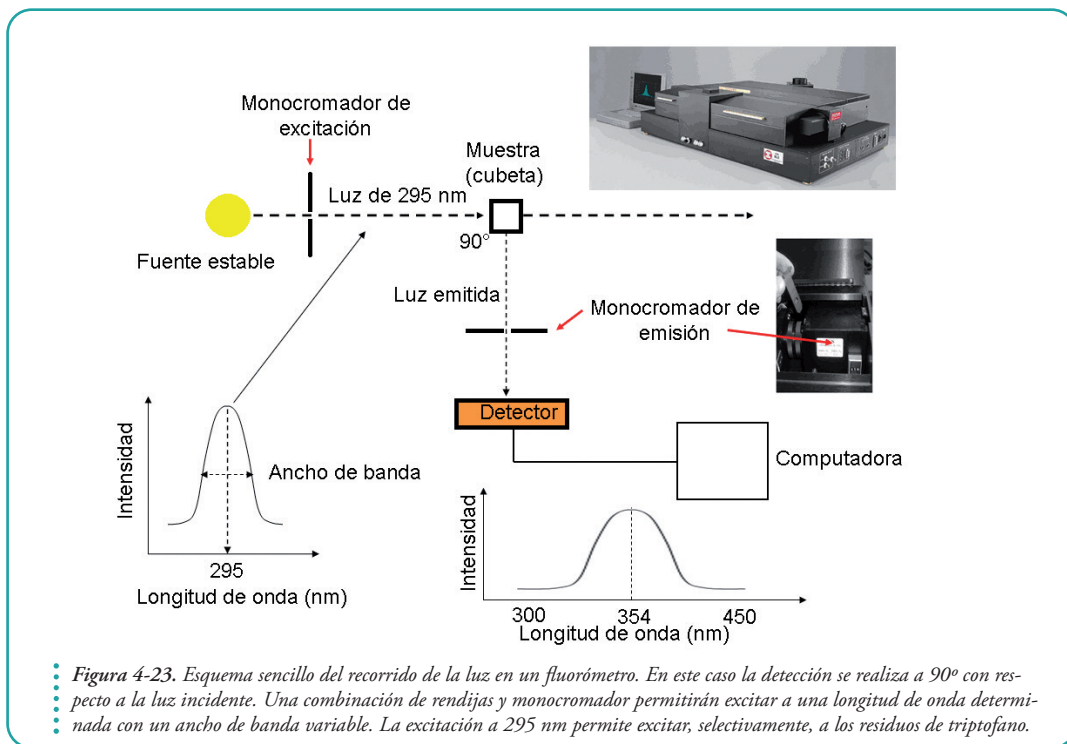


Tres proteínas distintas en las que un residuo de triptófano se encuentra en ambientes bien diferentes (color amarillo). En la azurina el triptófano está completamente oculto, emite fluorescencia con una longitud de onda extremadamente menor (308 nm) a la del triptófano disuelto en solvente acuoso (356 nm). En cambio, el triptófano del glucagón está expuesto al solvente y su emisión se produce a 350.8 nm; un caso intermedio es el de la proteína G en el que el triptófano emite a 342.5 nm de acuerdo con una exposición parcial al solvente. (ver Figura 4-22)

- Cambios conformacionales
- Asociación de subunidades
- Unión de sustratos
- Desnaturalización

Además la fluorescencia de triptófano es muy sensible a la de excitación por el choque con otras moléculas, tanto por grupos externos, por ejemplo por moléculas de agua, típicamente en el caso de triptófanos expuestos al solvente, como por grupos vecinos de la propia cadena polipeptídica. Por otro lado, los residuos de triptófano pueden ser excitados selectivamente a 295 nm, a esta longitud de onda los residuos de tirosina ya no absorben (ver absorción UV) y, así, se evita la excitación de las tirosinas.

Para poder excitar a una longitud de onda determinada y analizar la emisión en un rango deseado de longitudes de onda (espectro de emisión) los fluorómetros requieren ciertos componentes: un monocromador de excitación para fijar la longitud de onda de excitación y un monocromador de emisión para detectar fotones emitidos en las distintas longitudes de onda (ver **Figura 4-23**). Como se muestra en dicha figura, se excitará centrando la longitud de onda en un valor, pero con un ancho de banda que podrá ser variable (2, 4, 8 ó 16 nm por ejemplo).



Ahora sí, y volviendo a las proteínas, como hemos mencionado más arriba, el espectro de fluorescencia de triptofano variará según la accesibilidad al solvente de este residuo por lo menos por dos motivos:

- a. colisiones con moléculas de agua, que disminuirán la intensidad de fluorescencia porque las moléculas tendrán mayor probabilidad de desexcitarse por vías no radiativas y,
- b. relajación por solvente, en este caso lo que ocurre es un corrimiento de la longitud de onda de emisión hacia valores mayores (menos energéticos).

Además la fluorescencia de triptofano también variará acorde a la movilidad de la cadena polipeptídica, colisiones intra moleculares.

En la **Figura 4-24** se muestran espectros de fluorescencia de triptofano de la proteína β -lactamasa en tres estados conformacionales.

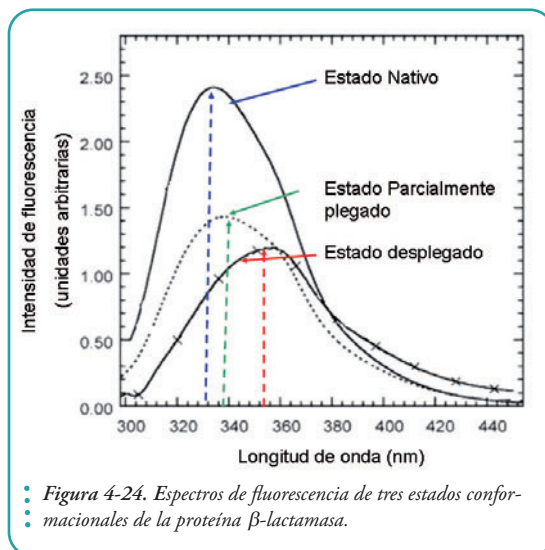
Como se puede ver en el estado nativo (estructura terciaria consolidada) las colisiones intra

e intermoleculares son mínimas (la intensidad de fluorescencia es alta, se minimizan las vías de desexcitación no radiativas por colisiones) y la emisión se produce desde un entorno relativamente apolar (335 nm). En cambio, en el estado, parcialmente, plegado las colisiones aumentan porque la cadena polipeptídica es mucho más flexible. Se puede ver que a pesar de esto, la emisión sigue siendo desde un entorno relativamente apolar (338 nm, no hay efecto por solvente). Cuando se observa el espectro del estado desplegado de esta proteína se detecta un corrimiento hacia longitudes de onda mayores 354 nm, que coincide con la longitud de onda de emisión del triptofano disuelto en agua (relajación por solvente). La intensidad es también baja, ahora también están presentes las colisiones con moléculas de solvente que desexcitan a una elevada proporción del fluoróforo por vías no radiativas.

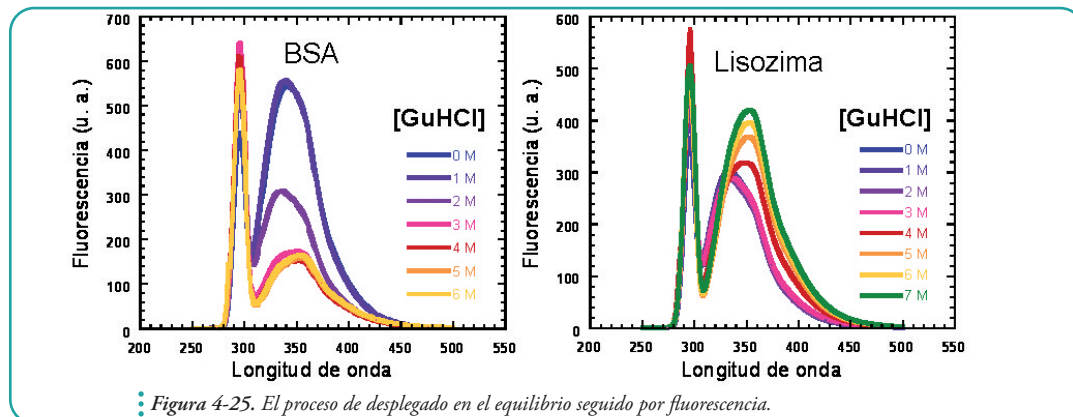
En la cadena polipeptídica existen algunos grupos químicos que pueden contribuir a la desexcitación por vías no radiativas. Entre ellos las histidinas y los puentes disulfuro son los más comunes.

En la **Figura 4-25** se muestra el proceso de desplegado en función de la concentración de desnaturalizante para dos proteínas: la BSA (albúmina bovina sérica) y la lisozima de clara de huevo.

En el caso de la BSA podemos ver que, a medida que se incrementa la concentración de desnaturalizante (GuHCl), la intensidad de fluorescencia disminuye por colisiones



En línea llena, el estado nativo; en línea punteada, un espectro correspondiente a una conformación parcialmente plegada y, en línea llena con cruces, el espectro correspondiente a la β -lactamasa desplegada. Se observan cambios en la intensidad del máximo de fluorescencia y en la longitud de onda relacionada al máximo consecuencia de cambios en el entorno químico de los residuos. (ver Figura 4-24.)



internas a bajas concentraciones de desnaturante y, luego, a mayores concentraciones de desnaturantes por colisiones con moléculas de agua.

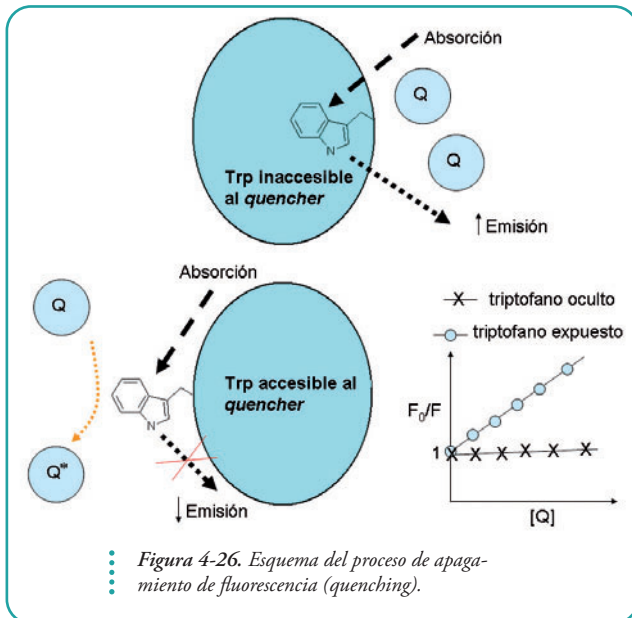
Por el contrario, en el caso de la lisozima la intensidad de fluorescencia aumenta a medida que se incrementa la concentración de GuHCl y que aumenta la concentración de proteína en estado desplegado. Esto se debe a la presencia de puentes disulfuro presentes en esta macromolécula. En el estado desplegado se incrementa la probabilidad de desexcitación con emisión de luz porque los residuos de triptofano están alejados de los puentes disulfuro.

Desnaturalización química de la proteína BSA y de la lisozima de clara de huevo de gallina, a la izquierda y a la derecha, respectivamente. Se excitó a 295 nm con un ancho de banda de 8,0 nm. El pico ubicado a 295 nm es producido por la existencia de dispersión luminosa debido a la presencia de partículas en la solución. (ver Figura 4-25)

Apagado y apagadores de la Fluorescencia

El retorno no radiativo al estado basal **por contacto con una segunda molécula** puede también ser explotado en el análisis conformacional de las proteínas. Este proceso es conocido con el nombre de *quenching* o apagado y, a estas moléculas, se las conoce como *quencher*s o apagadores.

En este tipo de experimentos se mide el espectro de fluorescencia con cantidades crecientes de apagador y se calcula F_0/F , donde F_0 es fluorescencia en ausencia de apagador y F es la



Arriba un triptofano oculto no interacciona con las moléculas apagadoras (quencher)s. Abajo, un triptofano expuesto puede deexcitarse por vías no radiativas ante la presencia en la solución de moléculas apagadoras de la fluorescencia. En el gráfico de la derecha se muestra este efecto: la reducción de la fluorescencia (y el aumento del cociente F_0/F , donde F_0 es la fluorescencia en ausencia de apagador) en función de la concentración de apagador (quencher). (ver Figura 4-26)

fluorescencia en presencia (ver gráfico en la **Figura 4-26**). A medida que agreguemos apagador, la fluorescencia será menor y el cociente F_0/F resultará en un valor mayor esto siempre y cuando el apagador tenga acceso al fluoróforo (curva que corresponde a triptofano ex-

puesto). Si no tuviera acceso al fluoróforo la pendiente de la recta debería ser cero (la intensidad de fluorescencia no debería variar por el agregado del apagador).

Estos gráficos se conocen como gráficos de Stern-Volmer y las rectas pueden representarse mediante la ecuación

Donde k_Q depende del tipo de fluoróforo y apagador, de la accesibilidad del fluoróforo. Además cuando el apagado depende de la difusión del apagador hacia el fluoróforo el proceso

$$\frac{F_0}{F} = 1 + k_Q \cdot [Q]$$

es dependiente de la viscosidad del solvente y de la temperatura.

Tres moléculas apagadoras de la fluorescencia de triptofano muy comúnmente usadas son el ión yoduro, la acrilamida y el gas **oxígeno** (O_2).

El hecho de que el O_2 penetre en el interior de las proteínas e incluso logre apagar a la fluorescencia de los triptofanos más ocultos **sirvió para demostrar que las proteínas son bastante más flexibles con respecto a lo que se creía en función de los datos cristalográficos.** Estos experimentos fueron algunos de los tantos experimentos geniales realizados por el científico de origen argentino Gregorio Weber (Figura 4-27) quien desarrolló una gigantesca batería de teoría y experimentos permitiendo comprender muchos de los procesos de fluorescencia en el campo del estudio de las proteínas.



Figura 4-27.
El Dr. Gregorio Weber.

Fluoróforos extrínsecos

Es muy importante destacar que existen innumerables moléculas fluorescentes muy útiles para el estudio de la estructura proteica. A estos fluoróforos, que no forman parte de la cadena polipeptídica (fluoróforos intrínsecos de la proteína), se los conoce con el nombre general de fluoróforos extrínsecos.

Este reactivo es capaz de unirse a las superficies hidrofóbicas accesibles presentes, típicamente, en los estados parcialmente plegados. Lo que resulta muy útil, en el caso de ANS, es que el rendimiento cuántico del ANS en solución es muy bajo pero, cuando se une a las superficies apolares proteicas, aumenta considerablemente (aumenta la intensidad de fluorescencia); además la excitación se realiza a 350 nm, y a esta longitud de onda las proteínas no absorben, por lo que se lo puede excitar selectivamente.

Este comportamiento permite determinar en la presencia de estos estados conformacionales en solución. Por otro lado el ANS libre (en solvente acuoso) emite a longitudes de onda cercanas a los 495-500 nm mientras que el ANS unido emite a 480 nm. La gran diferencia en las longitudes de onda de emisión ocurre porque el ANS que se ha unido a la superficie proteica se encuentra en entornos apolares; en cambio, en el caso del ANS libre existe relajación por solvente del estado excitado corriendo la longitud de onda hacia longitudes mayores, menos energéticas.

Un ejemplo es el 1,8-anilino naftaleno sulfonato (ANS, ver Figura 4-28).

Transferencia de energía por resonancia

Existe un proceso que es, particularmente, útil en el análisis de la conformación proteica y en el estudio del plegado de las proteínas. Es la transferencia de energía por resonancia. Este proceso involucra, por lo menos, a dos grupos químicos, uno llamado donador y otro que recibe el nombre de aceptor. Ambos pueden ser fluoróforos. El proceso consiste en la transferencia de la energía absorbida por parte del donador (y desexcitación del donador) al grupo aceptor sin emisión de luz. Si el grupo aceptor es un fluoróforo este puede, a su vez, desexcitarse emitiendo luz. (Figura 4-29).

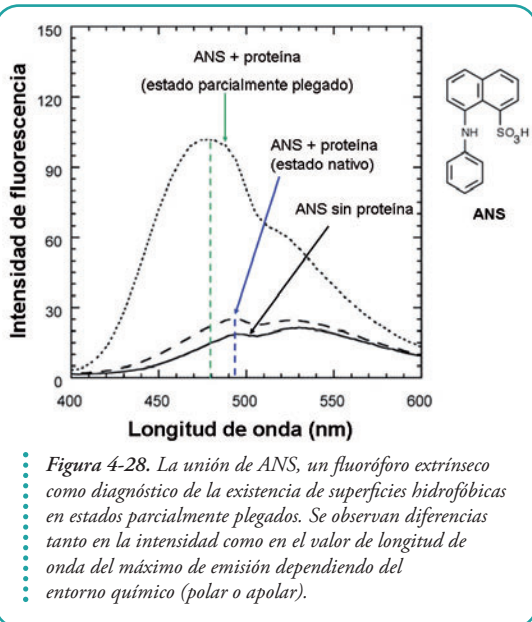


Figura 4-28. La unión de ANS, un fluoróforo extrínseco como diagnóstico de la existencia de superficies hidrofóbicas en estados parcialmente plegados. Se observan diferencias tanto en la intensidad como en el valor de longitud de onda del máximo de emisión dependiendo del entorno químico (polar o apolar).

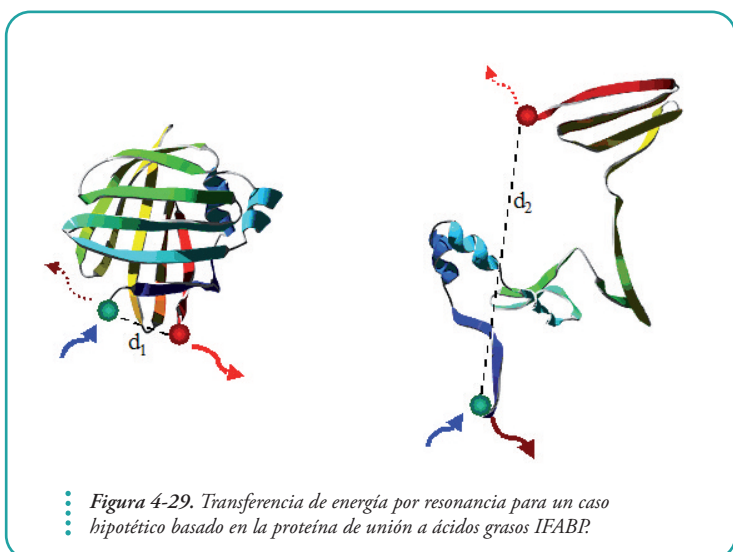


Figura 4-29. Transferencia de energía por resonancia para un caso hipotético basado en la proteína de unión a ácidos grasos IFABP.

En el estado nativo (izquierda), los fluoróforos ubicados en los extremos N y C-terminales quedan a una distancia óptima para la transferencia de energía, en cambio en el estado desplegado, en promedio, esta distancia es mucho mayor, impidiendo una transferencia eficiente. (ver Figura 4-29)

En la figura 4-29 se muestra, como ejemplo, un experimento hipotético en el que se han unido el fluoróforo donador a un extremo de la proteína (círculo verde) y el

fluoróforo aceptor al otro extremo (círculo rojo). Cuando la proteína está correctamente plegada ambos extremos se encuentran ubicados a una distancia d_1 que es óptima para la transferencia de energía por resonancia; al excitar al donador, este transfiere la energía (se desexcita) al grupo aceptor (quien se excita y luego se desexcita emitiendo luz) entonces la emisión de fluorescencia del grupo donador es mínima. Sin embargo,

cuando la proteína se despliega los extremos quedan ubicados en promedio a una distancia d_2 , mucho mayor). En este caso la transferencia de energía entre ambos fluoróforos será despreciable y la emisión por parte del grupo donador será máxima.

En rigor la distancia no es el único parámetro que determina la eficiencia de este proceso.

Además de la distancia entre donador y aceptor también juega un rol importante la orientación relativa entre ambos grupos y el tiempo que el grupo donador permanece en estado excitado antes de emitir un fotón. Por otro lado, es fundamental que exista solapamiento entre el espectro de emisión del donador y el espectro de absorción del aceptor (ver **Figura 4-30**).

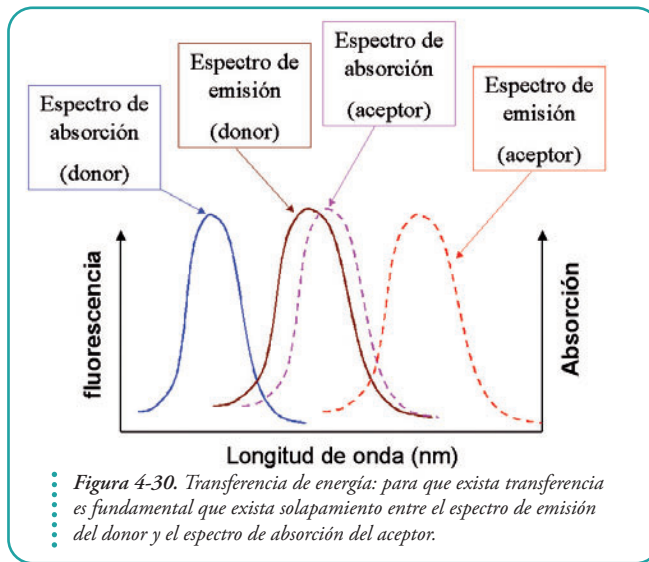


Figura 4-30. Transferencia de energía: para que exista transferencia es fundamental que exista solapamiento entre el espectro de emisión del donador y el espectro de absorción del aceptor.

La transferencia se puede medir observando la disminución de la fluorescencia del donador o el aumento de la emisión de aceptor pero, generalmente, se usa la primera forma.

La transferencia de energía por resonancia también es muy útil para detectar procesos en los que interviene la unión (por interacción) y el reconocimiento de dos moléculas.

Existen además diversos sustratos fluorescentes de enzimas en los que se detecta la ruptura de la molécula y la formación de los productos por transferencia

de energía por resonancia (**Figura 4-31**). Estos son, simplemente, dos ejemplos de las decenas de aplicaciones que existen en la actualidad.

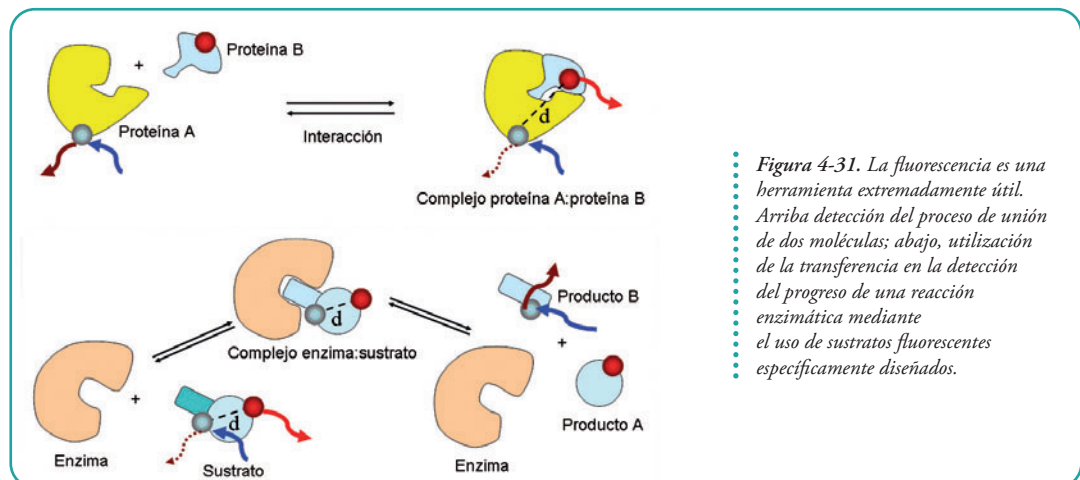


Figura 4-31. La fluorescencia es una herramienta extremadamente útil. Arriba detección del proceso de unión de dos moléculas; abajo, utilización de la transferencia en la detección del progreso de una reacción enzimática mediante el uso de sustratos fluorescentes específicamente diseñados.

Dicroísmo circular

Dicroísmo circular es la absorción diferencial por parte de una molécula asimétrica (en nuestro caso, de algunos grupos químicos de la cadena polipeptídica) de dos haces de luz circularmente polarizados de sentido opuesto (izquierdo y derecho). La asimetría de los cromóforos en proteínas (amidas, grupos aromáticos y puentes S-S) se induce por la interacción con grupos vecinos (el entorno químico).

Así, para comprender qué es dicroísmo circular hay que comprender algunas de las propiedades de la luz. Por ejemplo, alguna forma sencilla de representación de un haz de luz (Figura 4-32), qué es luz polarizada y qué es luz circularmente polarizada.

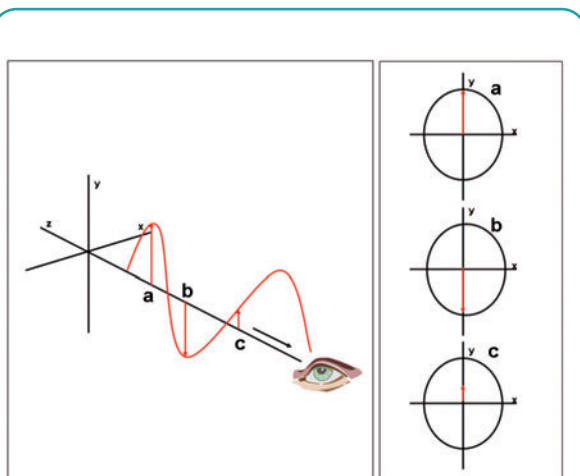
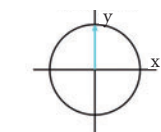


Figura 4-32. Dos sistemas de representación de un haz de luz.

Recordemos que podemos representar a la luz en función de su comportamiento ondulatorio. En el esquema de la izquierda un observador imaginario puede ver cómo avanza un haz en el espacio a través de la coordenada Z, mientras que si el observador imaginario se ubicara en el sitio indicado con la figura del ojo, siempre de frente al haz (luz linealmente o plano polarizada, ver más adelante), vería en este caso cómo cambia la amplitud del haz en función del tiempo. Esto está indicado por las flechas rojas en el panel de la derecha, por ejemplo, a tiempo $t=a$ la amplitud es máxima, a tiempo $t=b$ también es máxima en módulo pero su signo es opuesto, a tiempo $t=c$ la amplitud toma un valor intermedio. (ver Figura 4-32)

La luz blanca (la que emite una lamparita común y corriente es luz no polarizada: la fuente emite haces de luz en todas las direcciones (x, y, z) y sentidos. Sin embargo, existen elementos ópticos que permiten polarizar la luz, seleccionar los haces que vibran, por ejemplo en un plano, como en el caso de la Figura 4-32, en el que el haz vibra en el plano Z, Y.

Antes de atravesar la muestra ($t=a$)



Después de atravesar la muestra ($t=b$)

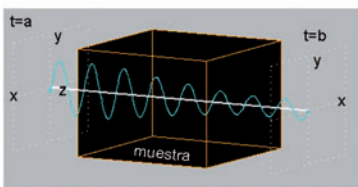
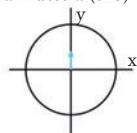


Figura 4-33. Representación de un medio capaz de absorber luz linealmente polarizada.

Si el haz de luz polarizada atravesara un medio (una muestra) que absorbe luz el observador imaginario ahora vería algo parecido a lo que se muestra en la Figura 4-33: la amplitud máxima del haz se va atenuando en función de su paso por la muestra.

Por otro lado, podemos pensar qué pasaría si sumáramos dos

haces de luz (en la Figura 4-34 se lo representa). El haz en rojo y el haz verde están en fase (desfasados en 0°), esto quiere decir que llegan a su máximo de amplitud simultáneamente. El haz representado en rojo está polarizado en el plano z, x ; mientras que el haz representado en verde está polarizado en el plano z, y ; es decir que ambos haces están polarizados en planos uno a 90° del otro. La resultante de la suma de ambos haces **también es luz polarizada en el plano** y se representa con trazo celeste. En los tiempos $t=2$ y $t=4$ la amplitud del haz es cero, mientras que en los tiempos $t=1$ y $t=3$ el módulo de la amplitud es máximo.

La amplitud máxima del haz se va atenuando en función de su paso por la muestra. (ver Figura 4-33)

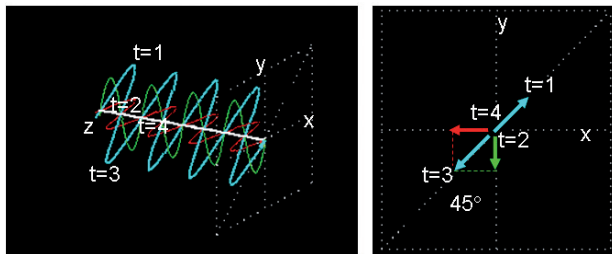


Figura 4-34. Suma de dos haces en fase de luz linealmente polarizada en planos perpendiculares (haces en color verde y rojo). El haz resultante se muestra en celeste; en este caso también es linealmente polarizado.

Sin embargo, si sumamos ahora dos haces de luz linealmente polarizada (como en el caso anterior, polarizados en planos a 90° pero, **en este caso los haces están desfasados 90°** , esto quiere decir que cuando un haz llega a un máximo de amplitud (haz verde en plano z, y) el segundo haz tiene amplitud 0 y, cuando el primero tenga amplitud 0, el segundo tendrá amplitud máxima; el haz

resultante ya no estará polarizado en el plano sino que se propagará helicoidalmente, describiendo una hélice y, si el observador lo viera avanzar desde el frente vería que su amplitud no varía y que el haz resultante describe un circunferencia en función del tiempo (t) (ver Figura 4-35, panel de la derecha), ¡por esto se la llama luz circularmente polarizada!

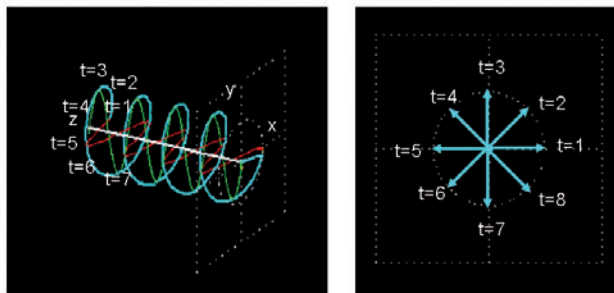
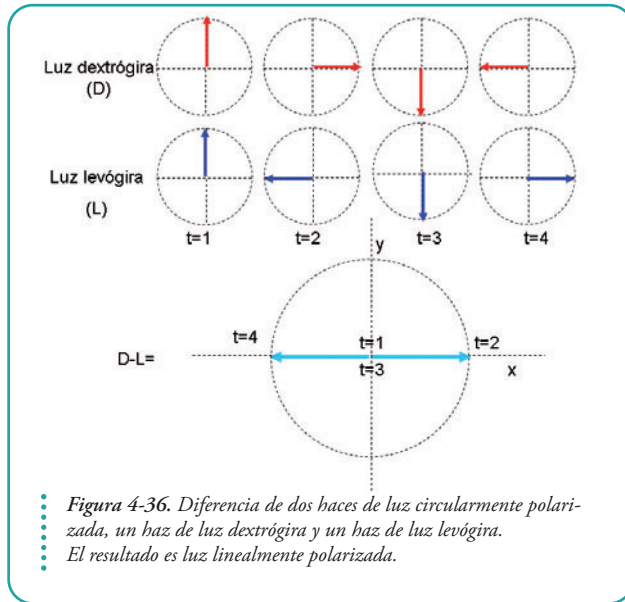


Figura 4-35. Suma de dos haces de luz linealmente polarizada, ambos polarizados en planos perpendiculares (a 90° , uno con respecto al otro). En este caso, además los haces están desfasados en 90° . Los tiempos $t=1, t=2, t=3$, etc. de ambas figuras se corresponden. El resultado es luz circularmente polarizada; en el espacio el haz se propaga helicoidalmente (izquierda). Dado que la suma a todo tiempo produce un vector de igual módulo un observador ubicado delante del haz podría ver cómo este describe un círculo.

Si la flecha celeste (el vector amplitud resultante) se mueve como las agujas del reloj, se dice que la luz es circularmente polarizada dextrógira (gira hacia la derecha), en cambio si gira hacia la izquierda (tomando como punto de referencia las 12 h en el reloj)

será luz levógira. Así pudimos fabricar luz circularmente polarizada a partir de dos haces linealmente polarizados.

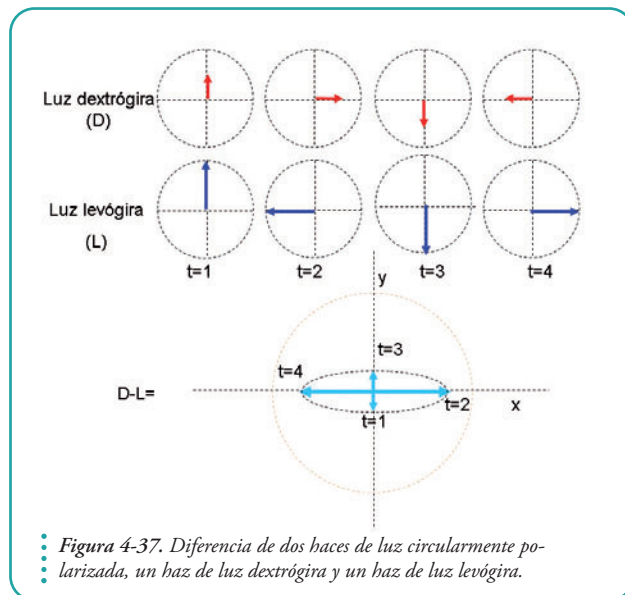
Por último podemos pensar qué es lo que pasaría si restáramos dos haces de luz, pero ahora en ambos casos los haces son de luz circularmente polarizada (Figura 4-36).



En la Figura 4-37 se muestra lo que ocurre: si hacemos la resta para cada tiempo (esto es restar la flecha roja a la azul a cada tiempo ($t=1, t=2, t=3$ y $t=4$, por ejemplo) recordando el signo (+, -) de cada cuadrante, Figura 4-38). Efectivamente, fabricamos luz linealmente polarizada, que vibra en el plano (x, z) y se propaga a través del eje z .

¿Qué pasaría si ambos haces atravesaran una muestra que absorbe diferencialmente luz circularmente polarizada?

Por ejemplo, la muestra absorbe únicamente luz circularmente polarizada dextrógira (no absorbe luz levógira).



En esta instancia, el observador imaginario ya no vería ni luz linealmente polarizada, ni tampoco circular. ¡Vería luz elípticamente polarizada! (el dibujo que describe el haz mirado de frente es una elipse, ver Figura 4-37), esto ocurre porque los haces no se cancelan completamente en la coordenada espacial y en nuestro ejemplo. Ya que como hubo absorción y, ahora, las amplitudes de ambos haces ya no son idénticas.

En general en muestras de proteínas la diferencia de absorción entre ambos haces es pequeña, por este motivo el radio menor de la elipse suele ser muy pequeño con respecto al mayor.



Uno de los problemas experimentales radica en detectar esta diferencia de absorción, es decir, detectar una diferencia en los coeficientes de absorción dextrógiro ε_D y levógiro ε_L . Se puede ver que el radio mayor de la elipse es proporcional a $\varepsilon_L + \varepsilon_D$ y que el radio menor será proporcional a $\varepsilon_L - \varepsilon_D$.

Podemos plantear dos ecuaciones de absorbancia, una para luz levógira y otra para luz dextrógira:

$$A_L = \log\left(\frac{I_{0L}}{I_{1L}}\right) \quad A_L = \varepsilon_L \times C \times l \quad A_D = \log\left(\frac{I_{0D}}{I_{1D}}\right)$$

Donde I_0 es la intensidad en ausencia de muestra (ausencia de absorción) e I_1 es la intensidad resultante luego de que el haz de luz atraviesa la muestra. Los subíndices L y D indican levógiro y dextrógiro, respectivamente.

Ahora podemos plantear la diferencia de absorbancias (usando Ley diferencial de Lambert-Beer)

$$A_L - A_D = \Delta A = \varepsilon_L \times C \times l - \varepsilon_D \times C \times l$$

Y en definitiva

$$\Delta A = (\varepsilon_L - \varepsilon_D) \times C \times l$$

Los equipos que permiten hacer estas mediciones (diferencias de absorción de luz circularmente polarizada dextrógira y levógira se llaman espectropolarímetros. Estos equipos son muy costosos y, electrónicamente, muy complejos; pero como veremos son extremadamente útiles. La relación entre los coeficientes de absorción puede extraerse a partir de la relación entre los radios de la elipse mediante propiedades trigonométricas:

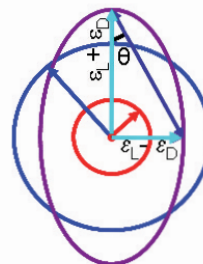
$$\operatorname{tg}(\theta) = \frac{\varepsilon_L - \varepsilon_D}{\varepsilon_L + \varepsilon_D}$$

Así una de las formas de medir absorción diferencial de luz circularmente polarizada (dextrógira y levógira) es medir el ángulo θ , en general el ángulo estará dado en miligrados. Éstas, que son las unidades crudas podrán relativizarse a la concentración de proteína, o enlaces peptídicos presentes en la muestra, también habrá que tener en cuenta el paso óptico (típicamente 0,1 ó 1 cm) de la cubeta que se utilizó. Se suelen usar las unidades de elipticidad molar $[\theta]$. En este caso se considera la concentración en deci moles ($1 \text{ dmol} = 0.1 \text{ mol}$) de enlaces peptídicos por litro (1000 cm^3).

Puede tenerse en cuenta el ejemplo numérico para más claridad

Una muestra que arroja una medición de 3 mili grados (0.003 grados) usando una cubeta de paso óptico de 1 cm y una concentración de proteína de $10 \mu\text{M}$ (0.00001 M). Si la proteína tiene una longitud de 101 re-

En este caso los haces han atravesado una muestra que absorbe, diferencialmente, luz circularmente polarizada en el sentido dextrógiro. El resultado es luz elípticamente polarizada. (ver Figura 4-37)



• Figura 4-38. Relación entre los coeficientes de absorción puede extraerse a partir de la relación entre los radios de la elipse mediante propiedades trigonométricas.

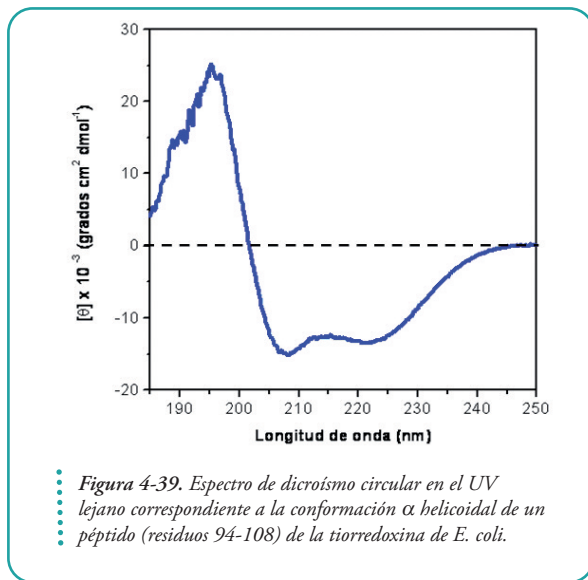
residuos de aminoácidos entonces tiene 100 enlaces peptídicos y la concentración de enlaces peptídicos va a ser 100 veces mayor que la de proteína (1 molécula tiene 100 enlaces peptídicos). Así la concentración de enlaces es $0.00001 \times 100 \text{ M} = 0.001 \text{ M}$ o 0.01 dM (decimolar)

$$[\theta] = \frac{0,003 \text{ grados}}{\frac{0,01 \text{ dmoldes}}{1000 \text{ cm}^3} \times 1 \text{ cm}} \quad [\theta] = 300 \frac{\text{grados}}{\text{dmoles}} \times \text{cm}^2$$

Volviendo ahora a nuestro punto, **las proteínas**, el dicroísmo circular (CD) en la zona UV lejano (180-250 nm), en la que absorbe el enlace peptídico, refleja el contenido de estructura secundaria y el CD en la zona UV cercano (250-340 nm), donde absorben las cadenas laterales de W, Y, F, H y los puentes disulfuro, revela características de la estructura terciaria a través de la detección de las asimetrías del entorno de esos residuos.

Los usos más comunes del dicroísmo circular en proteínas son: estimar el contenido de estructura secundaria, estimar el contenido de estructura terciaria, medir procesos de interacción de proteínas con otras moléculas (por ejemplo fármacos) y evaluar cambios conformacionales, y la estabilidad de la conformación a través de procesos de desnaturalización química o por temperatura.

Desde un punto de vista práctico vale la pena aprender a examinar espectros de dicroísmo circular porque, como podemos imaginar, son muy informativos.



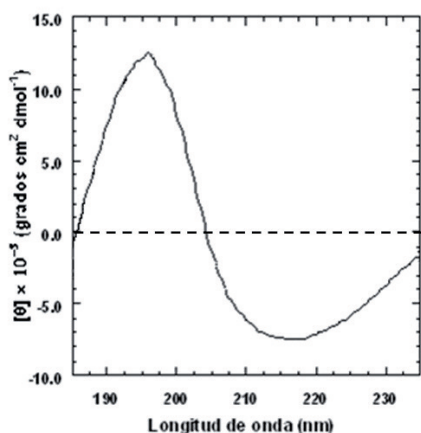
Para comenzar, en los espectros en la región del UV lejano (el enlace peptídico) podremos identificar bandas de absorción diferencial correspondientes a estructura α helicoidal: típicamente una banda positiva a 190-195 nm, y dos negativas una a 208 nm, y otra a 222 nm (ver Figura 4-39).

Las bandas de absorción diferencial correspondientes a estructura β están ubicadas a 195 nm (positiva) y a 215 nm (negativa) (ver Figura 4-40);

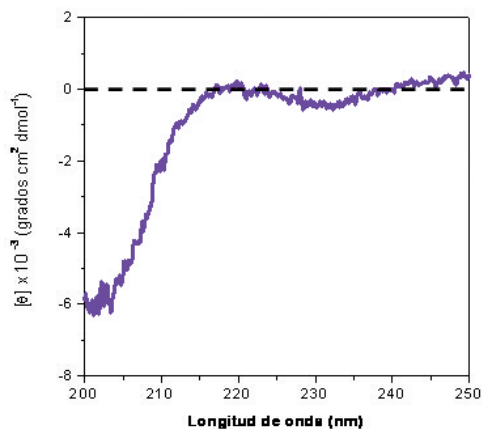
También es importante notar que el espectro de péptidos (y proteínas) desorganizados es, claramente, diferente a los anteriores. En la Figura 4-41 se muestra un ejemplo: el péptido TRX94-108 de la

tiorredoxina de *E. coli* (residuos 94-108) pero en su conformación desestructurada. Es importante notar que a 220 nm la banda negativa típica de elementos helicoidales ha desaparecido y a 190-200 nm tenemos una fuerte banda negativa.

Las diferencias observadas entre en los espectros permitirán estimar y, en el mejor de los casos, calcular el contenido de estructura secundaria. Así podremos *deconvolucionar* (descomponer el espectro en una suma de bandas de absorción) un espectro de una proteína con es-



• Figura 4-40. Espectro de dicroísmo circular en el UV lejano correspondiente a una proteína con estructura secundaria mayoritariamente β (IFABP, proteína de unión a ácidos grasos de intestino de rata)



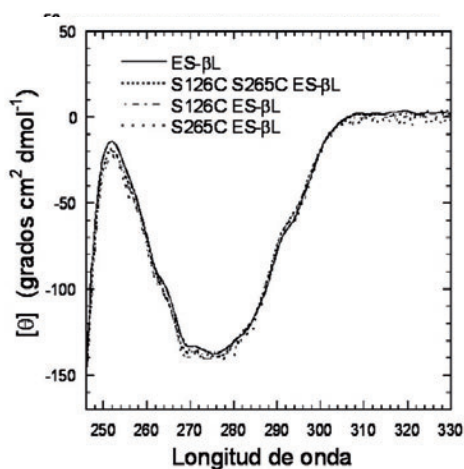
• Figura 4-41. Espectro de dicroísmo circular en el UV lejano correspondiente a un péptido desorganizado. Notar que a 220 nm la banda negativa típica de elementos helicoidales ha desaparecido y a 190-200 nm existe una fuerte banda negativa.

estructura compleja (por ejemplo $\alpha+\beta$) y estimar el porcentaje de estructura secundaria α y β y de regiones desorganizadas. Estas cuentas se hacen teniendo como base un conjunto de espectros de CD de proteínas con estructura secundaria conocida.

Para poder hacer esto, requerimos, entre otras cosas, conocer exactamente la concentración de proteína en la muestra medida al momento de la determinación, por ejemplo, por espectrofotometría UV (con un coeficiente de extinción confiable) porque tendremos que referir la absorción de luz circularmente polarizada (elipticidad) al contenido de enlaces peptídicos de la muestra.

Por último, el proceso de *deconvolución* de espectros no siempre es sencillo y las cuentas muchas veces se complican. Particularmente, es difícil la asignación de estructura β y vale la pena remarcar que algunas proteínas tienen espectros de dicroísmo circular en la región del UV lejano extremadamente atípicos.

La técnica de dicroísmo circular también permite estudiar el contenido de estructura terciaria. Los residuos aromáticos en las proteínas nativas se encuentran en ambientes asimétricos, propios de la estructura terciaria de las macromoléculas. Como se comentó con anterioridad, las bandas de absorción correspondientes a los residuos aromáticos se observan en el UV cercano (Figura 4-42).



• Figura 4-42. Espectros de dicroísmo circular en el UV cercano (240-340nm) de la β -lactamasa salvaje (ES- β L) y de tres variantes S126C (—•—), S265C (- - -), en estos dos casos se ha reemplazado un residuo de serina por uno de cisteína) y S126C S265C (- - -), un doble reemplazo de serinas por cisteínas.

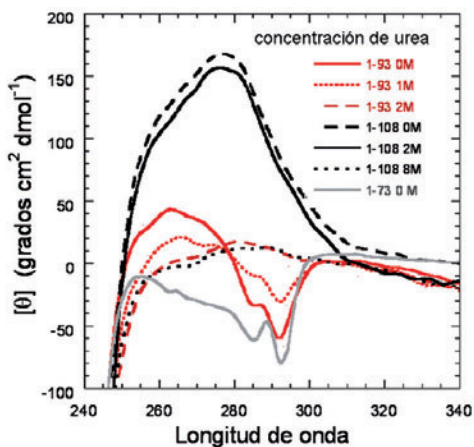


Figura 4-43. Espectros de dicroísmo circular en el UV cercano (240-340nm) de variantes de tioredoxina. En trazo color negro, la variante salvaje (longitud: 108 residuos), en ausencia (—) o en presencia de urea (8.0 M, (—) y 2.0 M (- - -)). Una variante corta (trazo rojo) TRX1-93 que comprende del residuos 1 al 93 en ausencia (—) o en presencia de urea 1.0 M (- - -) o 2.0 M (—). Por último la variante TRX1-73, una variante aún más corta, de 73 residuos en trazo gris.

Además puede verse la perturbación de los espectros en presencia de urea a causa de la desnaturalización y de la pérdida de estructura terciaria. En este caso no podemos decir que los entornos de los residuos aromáticos son similares para las tres variantes.

Por último, en la Figura 4-44 se muestra

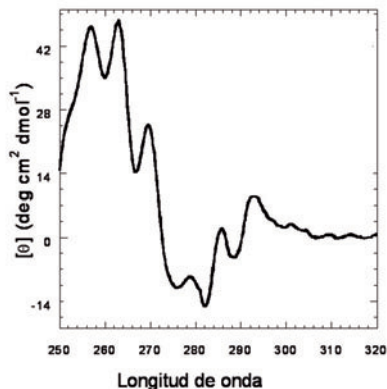


Figura 4-44. Espectro de dicroísmo circular en el UV cercano de la variante salvaje de IFABP (la proteína intestinal de unión a ácidos grasos de rata).

Lo realmente interesante es que el espectro de CD es una especie de huella dactilar del estado conformacional y en caso de que existan perturbaciones de la estructura terciaria podremos observarlas. Debemos tener presente que las bandas de absorción podrán ser positivas o negativas y que muchas veces unas (correspondientes a un residuo localizado en un entorno de la estructura) podrán cancelarse con otras correspondiente a otro residuo localizado en un ambiente distinto. Aquí se muestran tres ejemplos.

En la Figura 4-42 se pueden ver espectros de CD en el UV cercano de la proteína β -lactamasa de *Bacillus licheniformis* y de tres variantes mutantes dos variantes simples en las que se reemplazó un residuo de serina por uno de cisteína y una doble mutante en la que se reemplazaron dos residuos de serina por cisteínas. La superposición de los espectros muestra que el empaquetamiento terciario en el entorno de los residuos aromáticos sería semejante para las tres variantes con respecto a la proteína salvaje (sin mutaciones).

En la Figura 4-43 se muestran espectros correspondientes a tres variantes de la tioredoxina de *E. coli*: la proteína salvaje (1-108) y variantes acortadas (1-93) y (1-73)

un espectro de CD en el UV cercano de la variante salvaje de IFABP (la proteína intestinal de unión a ácidos grasos de rata). Podemos ver como se remarcan en este caso las bandas vibrónicas de los residuos de fenilalanina (entre 250 y 270 nm). Claramente este espectro es una huella digital de la presencia de estructura terciaria en la IFABP.

Finalmente vale la pena tener en cuenta algunas consideraciones referentes a aspectos instrumentales de estos equipos que utilizan fuentes de luz intensas, fotodetectores (detectores de fotones) eficientes y electrónica superior para reducir ruido. Con este objetivo (disminuir el ruido y aumentar

la señal) suelen tomarse algunos recaudos en la ejecución de los experimentos: por ejemplo se realizan muchos espectros que luego se promedian. Además se aumenta el tiempo de colección de datos. También suelen utilizarse celdas de 1 a 10 cm en la zona UV cercana; para poder medir señales débiles y celdas de 1, 0.5 y 0.1 mm en la zona UV lejana, para minimizar la absorción del solvente. Siempre se utiliza un flujo de gas N₂ continuo para evitar daño por ozono de la óptica (particularmente, de los espejos que forman parte del instrumento).

Espectrometría de masa.

La espectrometría de masa es, sin la menor duda, una de las herramientas más valiosas. Esta herramienta permite estudiar la estructura covalente de las proteínas y comprobar su integridad con gran rapidez y facilidad. Permite determinar la masa de una macromolécula con exactitud y gran precisión. Actualmente, dependiendo del tipo de equipamiento podemos realizar determinaciones en un rango increíblemente grande de masas. Podemos detectar muy rápidamente si la macromolécula posee o no modificaciones post-traduccionales o si su estado de oxidación es el correcto; si hemos producido una proteína recombinante y mediante técnicas de ingeniería genética hemos reemplazado alguna cadena lateral por algún otro aminoácido, podremos comprobar la modificación introducida. Es una técnica muy completa que se la aprovecha por su inmensa capacidad en la identificación (desde átomos a moléculas complejas). Permite identificar y cuantificar mezclas complejas; es rápida y de gran sensibilidad (entre otras cosas proporciona información estructural e información isotópica); es una técnica de rutina en procesos industriales y en el control de calidad de productos biotecnológicos; permite estudiar desde procesos de unión de pequeñas moléculas hasta el arreglo de estructura cuaternaria de complejos proteicos gigantes (> 800.000 Da); espectrometría de masa, no es una técnica espectroscópica porque no está basada en la interacción luz-materia.

Los espectrómetros de masa poseen cinco partes fundamentales (Figura 4-45):

1. un sistema de inyección de la muestra;
2. un sistema de ionización pasaje de la muestra a una fase de tipo gaseoso;
3. un analizador;
4. un detector de masas y
5. un sistema de captura y tratamiento de datos

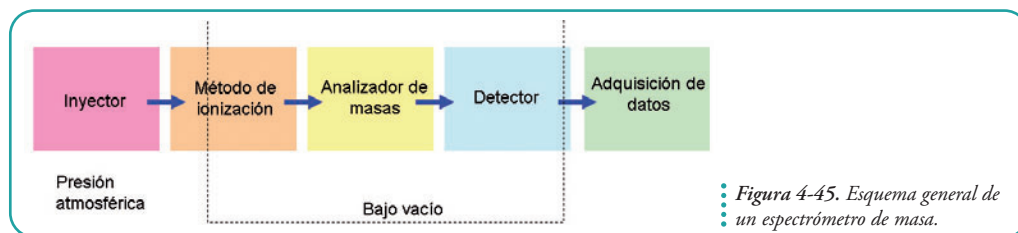
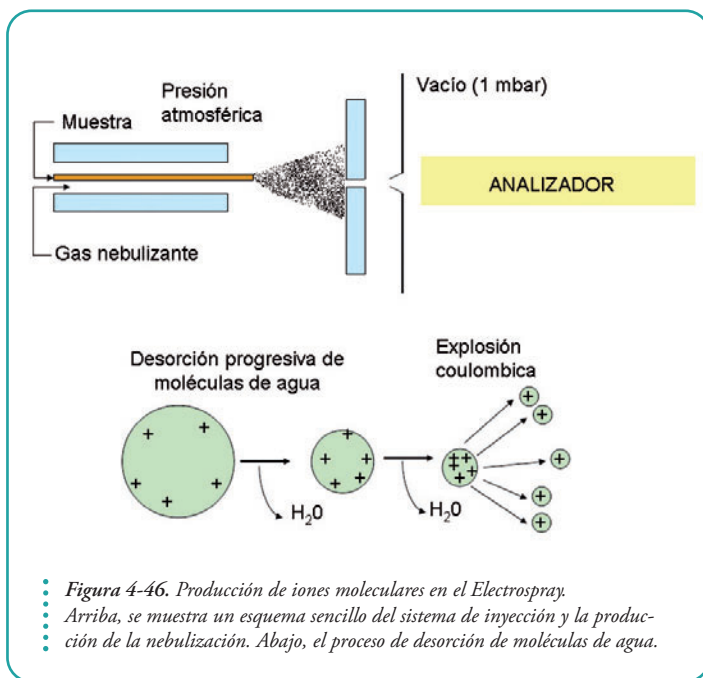


Figura 4-45. Esquema general de un espectrómetro de masa.

Existen dos o tres metodologías, ampliamente, aplicadas que permiten ionizar y, literalmente, *hacer volar a las proteínas* para su análisis. Sin duda una de las más impor-

tantes es ESI-MS: espectrometría de masa de ionización por *electrospray*. Esta técnica se basa en que la solución de proteína es inyectada a través de una micro aguja con propiedades especiales que permite la formación de un spray de la solución proteica (podemos pensar en una nebulización) en pequeñas gotas. Además, las moléculas son sometidas a un voltaje elevado y son conducidas mediante la aplicación de vacío. La desorción de moléculas de agua de la superficie de las gotas permite la generación de una relación tal, de cargas sobre la superficie de la gota, que su repulsión provoca explosiones coulombicas produciendo gotas más pequeñas y más pequeñas. Finalmente, para cada macromolécula cargada (ión molecular) sólo unas pocas moléculas de agua continúan adsorbidas a la superficie molecular (ver Figura 4-46).



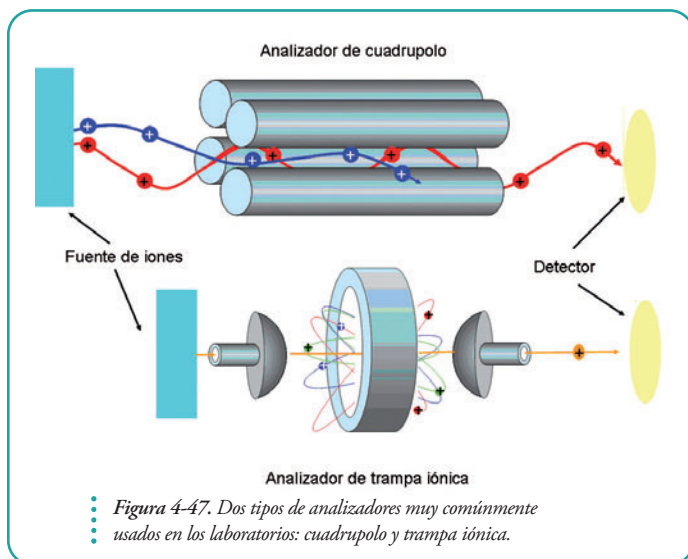
El proceso de evaporación es típicamente suplementado con un flujo de gas nitrógeno y calor.

Una característica importante de de ESI-MS es que, como las condiciones usadas para producir las gotas cargadas positivamente son ácidas, se tienden a protonar todos los sitios básicos en las moléculas de analito. Esto generará una familia de iones, algunos más cargados y otros menos cargados. Existirán entidades moleculares (iones moleculares) más probables y otras menos probables, dependiendo de la estructura primaria de la macromolécula.

El tercer componente del

equipo, como se mencionó más arriba, es el analizador. Los analizadores de masas llevan adelante la separación de los iones generados de acuerdo con su relación masa/carga (m/z). Existen disponibles tres o cuatro tipos de analizadores de masas distintos entre ellos el analizador cuadrupolo, la trampa de iones. En el caso del cuadrupolo y la trampa iónica se aprovecha el hecho de que frente a campos electromagnéticos, los distintos iones responderán de manera diferente. En particular, un analizador de cuadrupolo está compuesto por cuatro cilindros metálicos paralelos que forman un complejo campo electromagnético. Los iones pasarán a través del cuadrupolo (Figura 4-47) hacia el detector.

Las trayectoria de cada tipo de ión será afectada por el campo y sólo iones en los que la relación m/z tome valores determinados podrán atravesar el cuadrupolo, longitudinalmente, evitando colisionar con los cilindros. El resto de los iones chocarán con las superficies del analizador. Así, el cuadrupolo funciona como un filtro de iones. La trampa de iones también



utiliza campos electromagnéticos oscilantes pero, en este caso, para atrapar iones en forma controlada. Los iones son atrapados en un pequeño volumen por medio de campos eléctricos. Para ello, se dispone de un arreglo de tres electrodos, uno circular y dos electrodos hiperbólicos. La variedad de potenciales posibles que pueden aplicarse a los electrodos permite atrapar iones dentro de un rango m/z específico: atrapar iones con un valor m/z seleccionado, atrapar iones por encima de un valor m/z determinado, o

expulsar de la trampa iones de con valor m/z específico para su posterior detección.

No nos detendremos en análisis de los distintos tipos de detectores de iones, pero sí en el análisis de los datos experimentales. Si partimos de una muestra de proteína pura, por ejemplo de lisozima, a un pH determinado, observaremos una familia de iones que corresponderán a una única especie de proteína: lisozima. Estos iones difieren entre sí tanto en su carga, porque se han aceptado un número definido de cargas positivas provenientes de los protones (H^+) y, por otro lado, también difieren en su masa porque para cada ión molecular a la masa de la proteína deberemos sumarle la masa de los protones que ha aceptado. Por ejemplo, si una proteína de masa m y carga q acepta en su estructura 5 protones, su carga se modificará en $q+5$ y su masa en $m+5$ porque la masa de un protón es $m_H=1$.

La relación masa/carga en definitiva para unión estará dada por la ecuación:

$$\frac{m}{z} = \frac{m_p + n}{n}$$

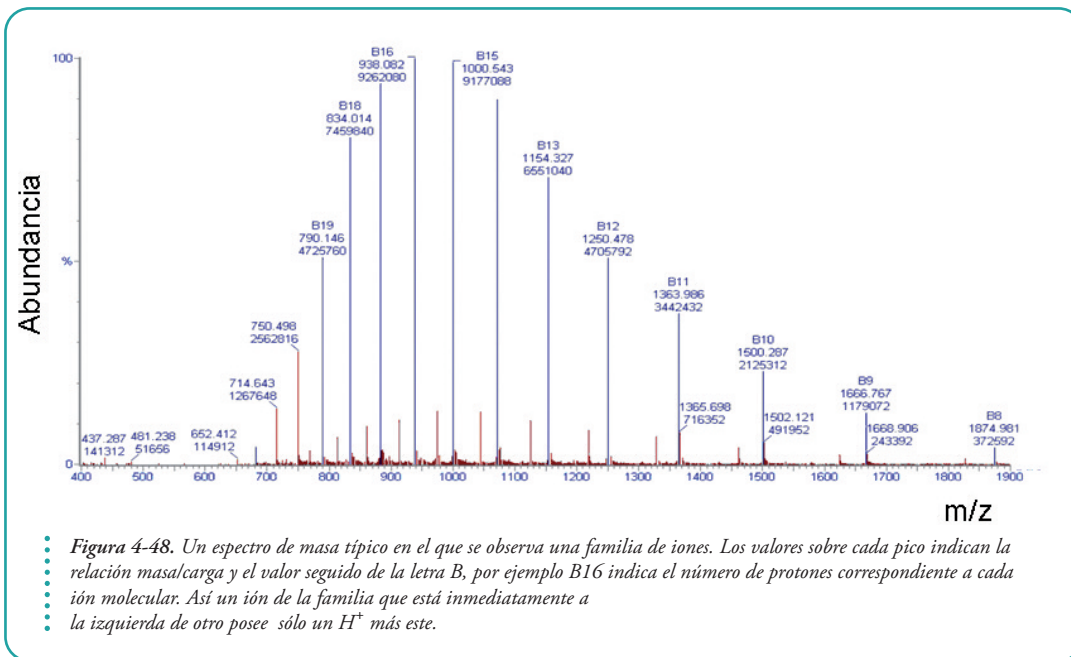
En las que m_p es la masa de la proteína, n es el número de protones que ha aceptado; si acepta un protón más entonces la ecuación se verá modificada de la siguiente manera

$$\frac{m}{z} = \frac{m_p + (n + 1)}{(n + 1)}$$

Dado que el espectrómetro de masa detecta iones en función de la relación masa/carga tendremos con dos iones un sistema de dos ecuaciones con dos incógnitas m y n . Sin embargo, en los espectros de masa muchas veces podremos observar muchos miembros de una familia de iones moleculares.

En la **Figura 4-48** se muestra un ejemplo. Se trata de un espectro de masa de la proteína de unión a ácidos grasos.

Es muy importante tener en cuenta que, dada la inmensa cantidad de átomos que posee una



proteína, debe tenerse en cuenta la abundancia de isótopos en la naturaleza. Por lo que **se usará en el cálculo una masa promedio** y no la masa del isótopo más abundante (monoisotópica).

Por ejemplo, la abundancia del hidrógeno cuya masa es de 1,007825 es del 99,99 % mientras que la masa del deuterio es de 2,014102 y su abundancia es del 0,015%. (ver **Tabla 4-2** y **Tabla 4-3**)

En este ejemplo la masa calculada experimentalmente es de 14.993,78±0,87 mientras que el valor teórico es de 14993. Queda más que claro con este ejemplo sencillo la relevancia que tiene este tipo de herramienta de extrema precisión en el estudio de proteínas.

El cálculo de la masa teórica puede hacerse usando el programa ProtParam disponible en <http://ca.expasy.org/cgi-bin/protparam>.

Nombre	Símbolo	Masa	Abund.	Símbolo	Masa	Abund.	Símbolo	Masa	Abund.
Hidrógeno	H(1)	1.007825	99.99	H(2)	2.014102	0.015			
Carbono	C(12)	12.000000	98.90	C(13)	13.003355	1.10			
Nitrógeno	N(14)	14.003074	99.63	N(15)	15.000109	0.37			
Oxígeno	O(16)	15.994915	99.76	O(17)	16.999131	0.038	O(18)	17.999159	0.20
Azufre	S(32)	31.972072	95.02	S(33)	32.971459	0.75	S(34)	33.967868	4.21

Tabla 4-2. Abundancias relativas de los distintos isótopos de los elementos de interés.

RESIDUO	CODIGO	COMPOSICION	MASA MONOISOTOPICA	MASA PROMEDIO
Glicina	G	C ₂ H ₃ NO	57.02146372	57.02146372
Alanina	A	C ₃ H ₅ NO	71.03711379	71.0779
Serina	S	C ₃ H ₅ NO ₂	87.03202841	87.0773
Prolina	P	C ₅ H ₇ NO	97.05276385	97.11518
Valina	V	C ₅ H ₉ NO	99.06841392	99.13106
Treonina	T	C ₄ H ₇ NO ₂	101.0476785	101.10338
Cisteína	C	C ₃ H ₅ NOS	103.0091845	103.1429
Leucina	L	C ₆ H ₁₁ NO	113.084064	113.15764
Isoleucina	I	C ₆ H ₁₁ NO	113.084064	113.15764
Aspargagina	N	C ₄ H ₆ N ₂ O ₂	114.0429274	114.10264
Acido Aspártico	D	C ₄ H ₅ NO ₃	115.026943	115.0874
Glutamina	Q	C ₅ H ₈ N ₂ O ₂	128.0585775	128.12922
Lisina	K	C ₆ H ₁₂ N ₂ O	128.094963	128.17228
Acido Glutámico	E	C ₅ H ₇ NO ₃	129.0425931	129.11398
Metionina	M	C ₅ H ₉ NOS	131.0404846	131.19606
Histidina	H	C ₆ H ₇ N ₃ O	137.0589119	137.13928
Fenilalanina	F	C ₉ H ₉ NO	147.0684139	147.17386
Arginina	R	C ₆ H ₁₂ N ₄ O	156.101111	156.18568
Tirosina	Y	C ₉ H ₉ NO ₂	163.0633285	163.17326
Triptofano	W	C ₁₁ H ₁₀ N ₂ O	186.079313	186.2099

ELEMENTO	CODIGO	MASA MONOISOTOPICA	MASA PROMEDIO
Hidrógeno	H	1.007825032	1.00794
Carbono	C	12	12.0107
Nitrógeno	N	14.00307401	14.0067
Oxígeno	O	15.99491462	15.9994
Azufre	S	31.97207069	32.065

Tabla 4-3. Masas correspondientes a los residuos de aminoácidos. Las masas de los elementos fueron extraídas del National Institute of Standards and Technology de Estados Unidos de Norteamérica. <http://physics.nist.gov/PhysRefData/Compositions/index.html>.

La obtención de estructuras tridimensionales por difracción de rayos X.

La determinación de la estructura tridimensional por difracción de rayos X requiere contar con cristales de proteínas.

¿Qué es un cristal de proteínas?

Un conjunto gigantesco de macromoléculas proteicas (unas 10¹⁵, mil millones de millones de moléculas) perfectamente ordenadas con una simetría determinada. Los cristales de proteína son muy espaciales, son frágiles porque el contenido de moléculas de agua es muy grande (típicamente mayor al 50%) y los contactos intermoleculares que mantienen a las proteínas empaquetadas en el cristal no son tantos ni tan fuertes como en el caso de otro tipo de cristales.

Los cristales crecen sólo en soluciones en las que la concentración de proteína debe ser inicialmente muy elevada (entre 5 y 20 mg/mL). A estas concentraciones, las proteínas suelen tener alta tendencia a agregar, pero en algunas condiciones particulares también cristalizan: forman agregados proteicos pero extremadamente ordenados.

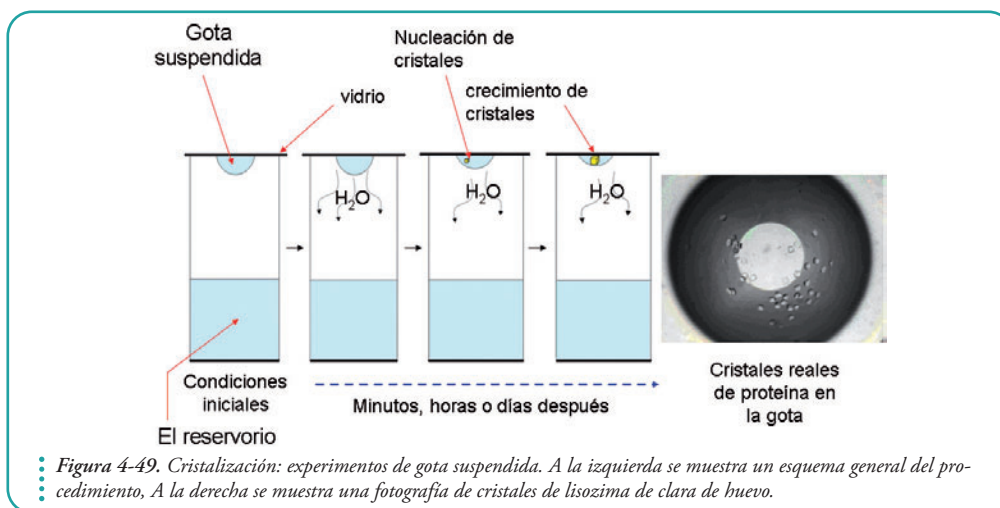
Uno de las dificultades experimentales que los cristalógrafos deben solucionar es, justamente, encontrar esas condiciones especiales de cristalización entre muchas no productivas (una aguja en un pajar). Típicamente, se hace la búsqueda de las condiciones de cristalización modificando

tipo y concentración de sales, cambios en pH.

Como en otros tipos de crecimientos cristalinicos debe formarse, primero, un núcleo de cristalización a partir del cual el cristal crecerá en todas sus dimensiones posibles. La presencia de impurezas (típicamente otras proteínas, o proteína mal plegada), partículas de polvo, etc. Frecuentemente, inhiben el crecimiento del cristal.

Existen muchas maneras de producir cristales de proteínas. Una de las formas más comunes es mediante experimentos de gota suspendida (Figura 4-49).

Por este motivo en la actualidad, en la mayor parte de los laboratorios, en los que se hace cristalogénesis, el proceso de generación de cristales de proteína, se usan sistemas automatizados, robots específicamente diseñados para mezclar las soluciones en micro gotas.



En este tipo de experimentos se prepara una solución de la proteína de interés, preferentemente, en agua libre de partículas e iones y se la diluye al medio con el *buffer* (por ejemplo, se agrega 1 μL de proteína más 1 μL de *buffer*) formando una pequeña gota que se deposita en un vidrio perfectamente limpio. Este vidrio se usa como tapa de un reservorio que tiene el mismo *buffer* que se usó para diluir la proteína (comúnmente unos 500 μL). Así, la concentración de los componentes del *buffer* en el reservorio será el doble que la concentración en la gota. El resultado es que por difusión de vapor, las moléculas de agua presentes en la gota difundirán hacia el reservorio, en principio, hasta que las concentraciones de la gota y del reservorio se equilibren. A medida que esto ocurre, la proteína, las sales, y el agente precipitante presente en el *buffer* de la gota, paulatinamente, comienzan a concentrarse. La velocidad de este proceso dependerá entre otras cosas de la temperatura. Por este motivo suele disponerse de cámaras de temperatura controlada (por ejemplo una 4 $^{\circ}\text{C}$ y al otra a 20 $^{\circ}\text{C}$).

A medida que la proteína y el agente precipitante se concentran comienzan a favorecerse las condiciones óptimas para la nucleación. Una vez que se forman los núcleos de cristalización, los cristales pueden continuar su crecimiento a expensas de la reducción de la concentración de proteína soluble en la gota.

Este tipo de dispositivos permiten probar, simultáneamente, muchas condiciones de

crystalización, sin consumir grandes cantidades de la proteína de interés, que por lo general es difícil de purificar.

Pero la obtención de cristales es, únicamente, una de las tareas de un cristalógrafo. Una vez obtenidos los cristales podrán realizarse los experimentos para obtener datos de difracción de rayos X.

¿Qué es la difracción de rayos X?

Podemos ir de a pasos pequeños y preguntarnos

¿Qué son los rayos X?

Los rayos X son radiaciones electromagnéticas de longitudes de onda extremadamente cortas, altamente energéticas. Estas radiaciones (fotones) son emitidos cuando los electrones se aceleran. Los haces de rayos X más poderosos se producen en un tipo de fuente conocida con el nombre de luz Sincrotrón.

Los sincrotrones son sistemas de aceleración de partículas en este caso electrones. Este tipo de haces son policromáticos, de espectro continuo, y permiten seleccionar una variedad de longitudes de onda entre ellas en el rango de 0,2 y 2 Å (ver más adelante, **Figura 4-51**).

Cuando los rayos X chocan con un átomo interactúan con sus electrones. Estos reemiten la radiación electromagnética incidente, pero, ahora, en diferentes direcciones. La frecuencia de la radiación reemitida es similar a la de la radiación incidente. Los fotones reemitidos por átomos cercanos interfieren entre sí constructiva o destructivamente, sumándose o cancelándose, porque las ondas ya no se encuentran necesariamente en las fases. Este es el fenómeno conocido con el nombre de difracción. En el ejemplo de la **Figura 4-49**, se muestra la consecuencia del fenómeno de difracción para dos objetos actuando como dispersores.

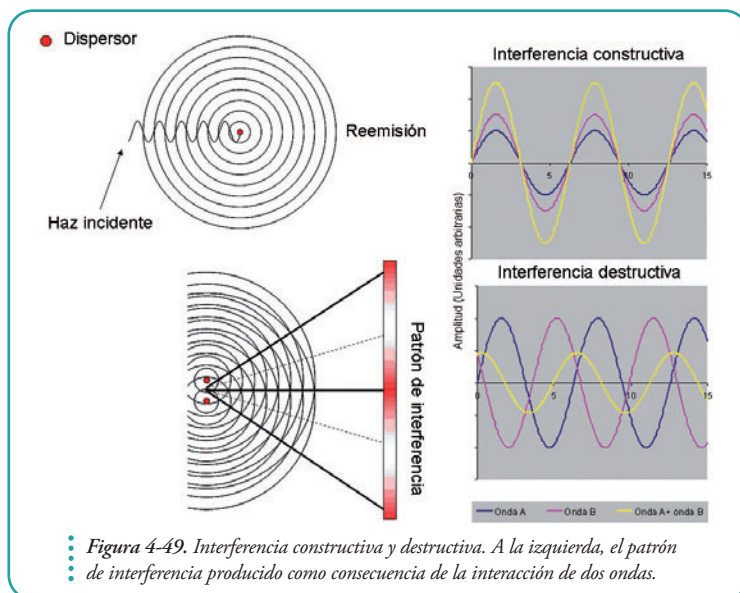


Figura 4-49. Interferencia constructiva y destructiva. A la izquierda, el patrón de interferencia producido como consecuencia de la interacción de dos ondas.

Así se genera un patrón de interferencia que es, en cada punto, igual a la suma (vectorial) de las ondas dispersadas. Podemos pensar como analogía en las ondas producidas al tirar una o dos piedras en una pileta.

Como la dispersión es el resultado directo de las interacciones entre los rayos X incidentes y los electrones de los átomos dentro del cristal, si logramos estudiar las direcciones

en las que ocurre difracción luego de que los rayos X impacten sobre el cristal, examinando la interferencia constructiva y destructiva, podríamos predecir la ubicación periódica de los átomos que lo componen (específicamente podríamos construir un mapa de las densidades de electrones en la red cristalina).

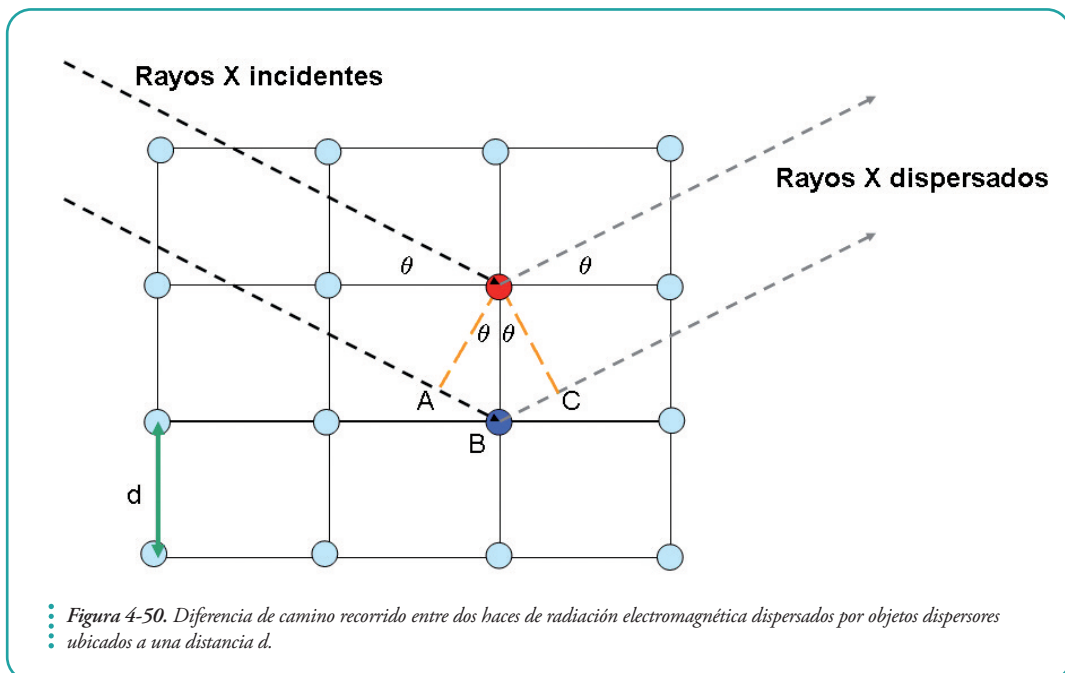
Gracias a los físicos William Henry Bragg y William Lawrence Bragg (su hijo) quienes derivaron una ley ahora conocida como Ley de Bragg podemos establecer relaciones entre los mapas de difracción y la estructura del cristal, es decir, la correspondencia entre la posición de los objetos que actúan como dispersores, y el patrón de interferencias (máximos y mínimos en intensidades). Los Bragg fueron premiados en el año 1915 con el Premio Nobel de Física por sus trabajos en la determinación de la estructura cristalina.

Efectivamente, descubrieron que los picos de intensidad en el patrón de difracción (máximos) para los rayos X dispersados podían ser seleccionados cuando el ángulo θ de incidencia era igual al de la radiación dispersada y, simultáneamente, la diferencia de camino realizada por la radiación era un número entero de veces el valor de la longitud de onda, de tal manera que se recuperaba la fase de la onda.

Como se ve en la **Figura 4-50**

$$\frac{\text{diferencia de caminos entre haces}}{2} = AB = BC = d \times \text{sen}\theta$$

$$\text{diferencia de caminos entre haces} = 2d \times \text{sen}\theta$$



Por esto, la intensidad debe ser máxima cuando

$$n\lambda = 2d \times \sin\theta$$

Donde n es un número entero y λ es la longitud de onda del haz de rayos X.

En este contexto, vale la pena preguntarse cómo están organizadas las proteínas dentro de un cristal (recordemos que son unas 10^{15} moléculas). Es muy importante tener en cuenta que en un cristal coexisten miles de macromoléculas, perfectamente, ordenadas y alineadas y con movilidad extremadamente restringida. El efecto del orden es el aumento de la intensidad de las señales dispersadas.

Un cristal puede dividirse en unidades mínimas (Figura 4-51). Cada unidad mínima recibe el nombre de **celda unidad** y cada una de ellas contiene toda la información necesaria para reconstruir el cristal en base, simplemente a la traslación de la celda unidad en las tres dimensiones. A su vez, la celda unidad puede tener simetría interna. En este caso se define la unidad asimétrica.

El cristal es montado de tal forma que se lo puede rotar, paulatinamente, con una increíble precisión y los datos de los distintos planos de difracción son adquiridos con cámaras especiales. Además, es importante mencionar que existen estrategias de adquisición de datos experimentales que garantizan que el investigador pueda contar con varios juegos completos de datos, es decir, de las posiciones e intensidades de los picos de difracciones del cristal necesarias, en principio, para construir un modelo experimental en el que se reconstruyen las posiciones de los átomos que dieron origen a la difracción.

En una celda unidad pueden alojarse más de una macromolécula. No necesariamente esto indicará que las moléculas tienen estructura cuaternaria en solución. Muchas veces se trata de dímeros cristalográficos, donde las interacciones intermoleculares están favorecidas por el empaquetamiento del cristal. Éste, por ejemplo, es el caso de la tiorredoxina de E. coli, (TRX), estructura PDB: 2TRX, en la que existen dos moléculas de TRX por celda unidad y sin embargo la proteína en solución es básicamente monomérica.

Molécula

Celda unidad

Cristal

Fuente de rayos X

Cristal montado

Mapa de difracción

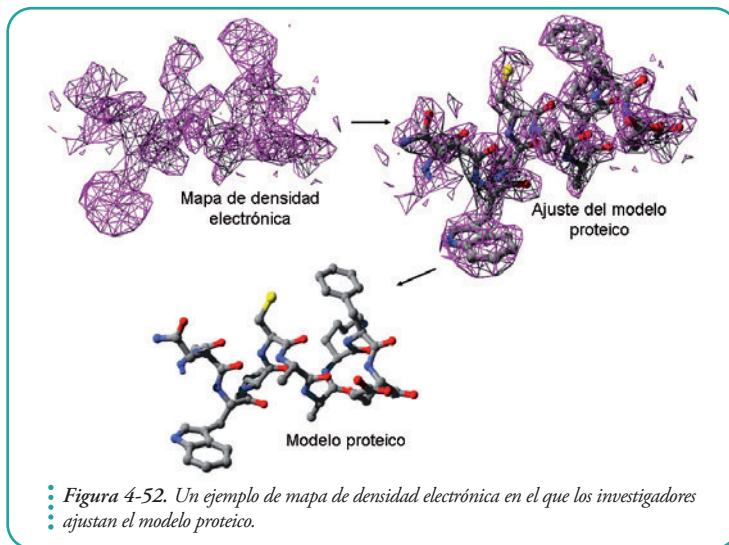
El Sincrotrón

Mapa de difracción real (un plano)

- Figura 4-51. El proceso para la colección de datos experimentales de difracción de rayos X es complejo e involucra elevada tecnología. La obtención de la proteína pura y correctamente plegada, su cristalización, la difracción y el análisis de los datos experimentales permitirá la subsiguiente construcción de un modelo. Se muestra una fotografía del Sincrotrón brasileño.
- <http://www.lnls.br/lnls/cgi/cgi.lua.exe/sys/start.htm?tpl=home>.

Utilizando matemática, relativamente avanzada, mediante transformaciones (transformaciones de Fourier) que relacionan densidad electrónica en el espacio con el patrón de difracción, se ubican en el cristal las densidades electrónicas que dieron origen a los picos de difracción de rayos X.

Para generar el modelo de la macromolécula el científico deberá hacer encajar (ajustar) el modelo molecular en las densidades electrónicas halladas experimentalmente por difracción (ver Figura 4-52). El mapa de densidad electrónica es una excelente aproximación de la posición de los átomos en de la macromolécula.



El investigador tendrá a su disposición, para hacer el ajuste del modelo a la densidad electrónica, la secuencia exacta de aminoácidos de la proteína, una biblioteca con las estructuras químicas de los veinte aminoácidos y de las geometrías moleculares más comunes, por ejemplo la de los enlaces peptídicos. El encaje o ajuste del modelo se hará básicamente modificando ángulos, longitudes de enlace y ángulos diedros (torsionales). Las

modificaciones tendrán un costo energético y deberán satisfacer simultáneamente

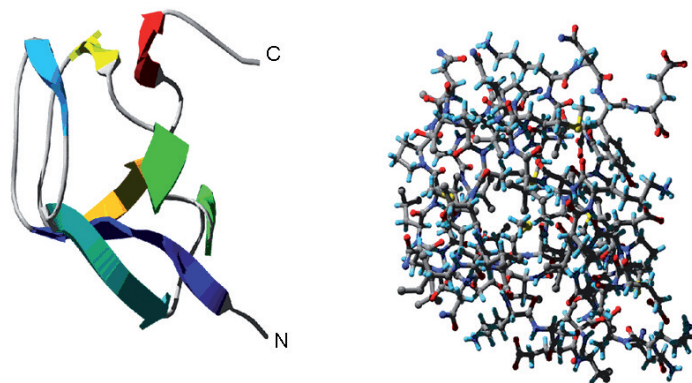
- a. El ajuste de la macromolécula a la densidades electrónicas
- b. Geometrías moleculares de menor energía

Así que, por lo general, los modelos cristalográficos se generan en función de las densidades electrónicas de los átomos de carbono (6 electrones) nitrógeno (7 electrones), oxígeno (8 electrones) y azufre (16 electrones) conocidos entre los cristalógrafos como átomos pesados.

En definitiva, un patrón de difracción proviene de una de la densidad electrónica y esta describe la estructura 3D de la macromolécula.

Cada *spot* en el patrón de difracción, a su vez, está determinado por una onda discreta de radiación electromagnética. Cada onda cuenta con una intensidad y una fase que describe su comportamiento. Al analizar el patrón de difracción vemos que lo único que podemos rescatar es una relación entre las intensidades

Vale la pena resaltar que ciertos átomos ricos en electrones como el azufre y algunos metales como el hierro y el cobre serán muy efectivos como dispersores de rayos X, mientras que por el contrario, los átomos pobres en electrones, como los átomos de hidrógeno que comparten su único electrón en un enlace no serán buenos dispersores, al punto que, por lo general, la densidad electrónica asociada a los átomos de hidrógeno no podrá ser observada. Solo será posible en experimentos en los que la resolución es máxima, por lo general, por debajo de 1,0 Å (ver Figura 4-53).

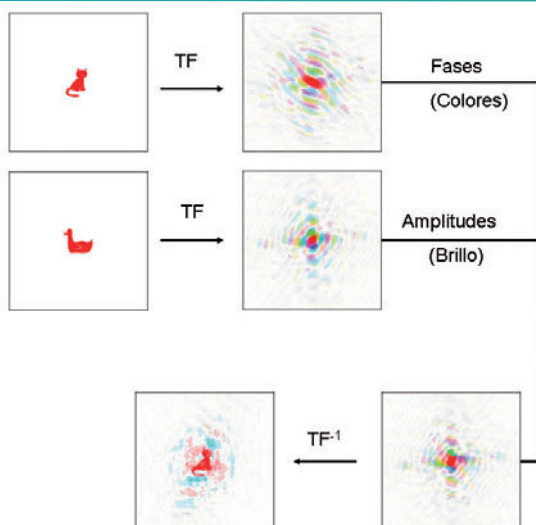


Proteína anticongelante de tipo III
PDB: 1UCS

• *Figura 4-53.* La estructura PDB: 1UCS, que corresponde a una proteína de la familia de las proteínas anti congelantes del grupo III, presentes en peces que sobreviven a temperaturas extremadamente bajas sin congelarse. Esta estructura posee una resolución increíblemente elevada de 0.65Å. A la izquierda una representación de cintas; en el panel de la derecha puede verse que la elevada resolución permite en este tipo de casos ubicar densidades electrónicas correspondientes a los átomos de hidrógeno.

de cada spot, pero en el camino hemos perdido las fases. Para reconstruir las densidades electrónicas debemos contar con ambas: con las intensidades y con las fases. El problema es encontrar la fase que corresponde a cada spot de difracción. Si bien no profundizaremos sobre este punto debe quedar claro que este es el principal problema con que se encuentran los cristalógrafos al punto que se puede ver que contar las fases adecuadas es mucho más importante que contar con las intensidades, las fases proveen la mayor parte de la estructura.

En la **Figura 4-54** se muestra con un ejemplo el efecto de utilizar las fases correctas e intensidades incorrectas para reconstruir una imagen.



• *Figura 4-55.* La importancia de las fases en cuanto a la información que éstas portan para la generación del modelo, en este caso una imagen. Se muestra el patrón de difracción o la transformación de Fourier (TF, que es una operación matemática equivalente) de la imagen de un pato y de la imagen de un gato. Podemos, entonces, utilizar el juego de fases exactas que dieron origen al patrón de difracción del gato y el juego de amplitudes exactas que dieron origen la difracción del pato. Si ahora hacemos la operación reversa, (creamos la imagen con intensidades de pato y fases de gato) obtenemos algo muy parecido al gato. En el patrón de difracción, los colores muestran las diferentes fases y el brillo del color indica intensidad. Reproducido por cortesía de Kevin Cowtan.

A esta metodología se la llamó **reemplazo isomórfico**, porque requiere que uno o unos pocos átomos pesados (cobre, uranio, oro, mercurio) estén específicamente ubicados en la unidad asimétrica; la presencia de estos átomos alterará el patrón de difracción modificando las intensidades de los haces dispersados pero sin afectar la estructura de la macromolécula ni su orientación en la celda unidad; sin afectar tampoco la estructura del cristal. Se deben adquirir datos de difracción del cristal nativo (sin átomos pesados) y del cristal con átomos pesados, que pueden incorporarse al mismo por difusión.

Este procedimiento permite ubicar los átomos pesados (mercurio o uranio) en la celda unidad y, como son sólo unos pocos átomos, se pueden usar estrategias que no se requieren información de las fases. Una vez que se ha determinado la ubicación de estos átomos, se pueden calcular las fases y las intensidades.

La contribución a la difracción de la proteína dependerá tanto de las fases que provienen de la proteína como de la fase de los átomos pesados; las intensidades se suman si ambos tienen la misma fase y se restan si tienen fases opuestas.

Si se conocen las fases de los átomos pesados puede, finalmente, estimarse las fases correspondientes a cada spot de difracción de la proteína.

Como de este procedimiento se obtienen 2 soluciones posibles por cada spot se realizan experimentos simultáneos con distintos metales. Así una única de estas dos soluciones será la real.

Existen otras metodologías que, en la actualidad, se utilizan más comúnmente. Muy resumidamente, las técnicas de reemplazo molecular usan modelos de estructuras homólogas de los que pueden calcularse fases esperadas: se desprenden fases iniciales que, luego, se optimizan en función de ajustes que tienen en cuenta el patrón de difracción obtenido experimentalmente. Otra metodología, ampliamente, usada es la dispersión anómala multi longitud de onda (MAD), requiere que los datos sean colectados a varias longitudes de onda, típicamente a tres longitudes, con el objetivo de evaluar efectos anómalos de absorción por parte de ciertos tipos de átomos. Estos efectos anómalos de absorción provocan diferencias de intensidades relativas para los spot de difracción de un único cristal al ser difractado con distintas longitudes de onda. Así puede calcularse la subestructura de átomos que producen MAD y, posteriormente, utilizar esas fases para calcular el resto. Una estrategia muy común para incorporar átomos que produzcan MAD y que estén ubicados selectivamente en el cristal es la incorporación de residuos de selenio metionina en lugar de metionina (se reemplaza el azufre por selenio).

Actualmente, uno de los mayores desafíos científicos consiste en estudiar estructuras de complejos multiprotéicos que van desde maquinarias de síntesis que incluyen polimerasas y ribosomas completos hasta virus. Sin duda comprender las bases de la función biológica a través del análisis estructural es parte de este gran desafío.

Si bien, actualmente, existen varias soluciones, la primera, históricamente, que responde al problema de las fases, se implementó para cristales de hemoglobina y fue desarrollada por Max Perutz y John Kendrew en el año 1954 (quienes recibieron el premio Nobel de Química en el año 1962 por estos descubrimientos).

Conclusión

Quinta Parte

Resulta fascinante pensar en la posibilidad de diseñar proteínas con actividades biológicas nuevas. Esto podría hacerse usando, como andamios, el esqueleto de proteínas conocidas y, más ambiciosamente, diseñando proteínas *de novo*.

En ambos casos, como buenos diseñadores o ingenieros necesitaríamos conocer las bases de la estabilidad proteica desde un punto de vista estructural y, también, las bases químicas y físicas de la función biológica que queremos que la proteína en cuestión cumpla. En el segundo caso, necesitamos, además, conocer el código que permite predecir una estructura a partir de una secuencia de aminoácidos determinada.

De cualquiera de las dos maneras será un requisito básico para el diseñador conocer la estructura tridimensional para realizar la ingeniería de la nueva función. Los grupos químicos que la desempeñen deberán estar ubicados y orientados en el espacio con una precisión absoluta.

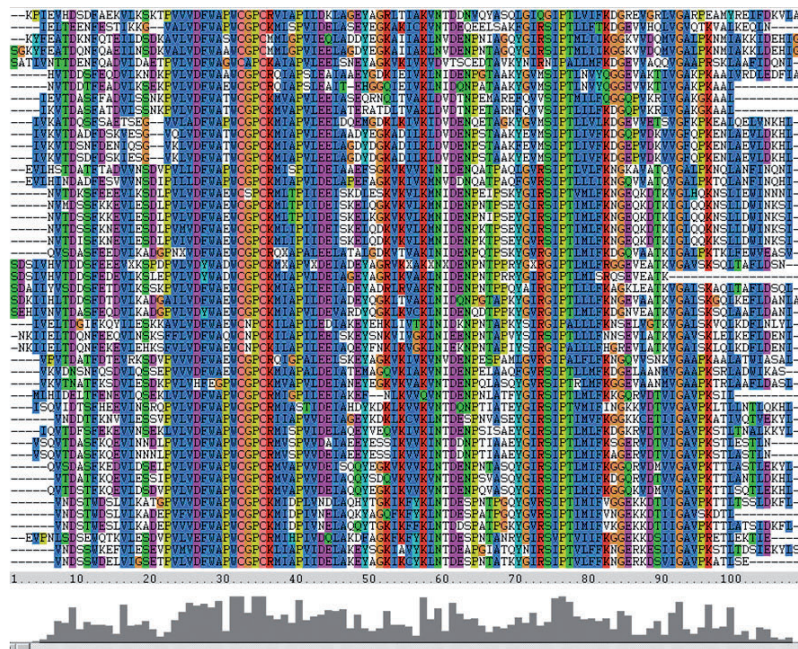
Uno de los problemas más importantes que los científicos intentan resolver en la actualidad es, justamente, comprender cómo la secuencia de aminoácidos en última instancia determina la estructura proteica, la búsqueda de un código para el plegado. Por ahora este problema no tiene una solución formal, sólo contamos con respuestas insipientes.

Sin embargo, podemos mencionar algunas generalidades. Hemos visto que existen familias de proteínas y que, los miembros de estas familias, tienen secuencias homólogas. Estos homólogos (que cuentan con un porcentaje de la secuencias conservada) tienen similitud en sus secuencias, similitud en su estructura y muchas veces en su función. Por lo general los residuos que suelen conservarse entre proteínas homólogas son residuos del corazón de la proteína pero, también, se conservan los relacionados con la función,

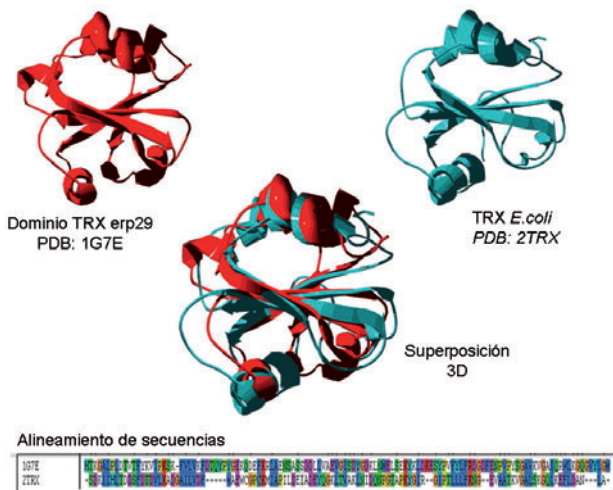
si es que estos homólogos cumplen similar función biológica (ver **Figura 5-1**)

En muchos casos, sin embargo, un mismo plegado puede estar asociado con funciones proteicas bien distintas (**Figura 5-2**).

Por esto, es muy riesgoso extraer conclusiones apresuradas acerca de los motivos por los cuales un residuo se ha conservado a lo largo de la evolución biológica de una secuencia.



• **Figura 5-1.** Conservación del sitio activo WCGPC en las tioredoxinas. Para este ejemplo se utilizó la base de datos de secuencias homólogas ConSurf. Se obtuvieron 50 homólogos. Puede verse que coexisten en la secuencia regiones menos conservadas, particularmente, en los extremos N y C-terminales y que ciertos bloques de residuos apolares se conservan (por ejemplo aminoácidos 22-25 y 78- 81); estos residuos apolares corresponden a residuos del corazón de la proteína. El alineamiento de estructuras primarias fue realizado con el programa de libre acceso ClustalX.



• **Figura 5-2.** Comparación de las estructuras primaria y terciaria de dos proteínas homólogas, evolutivamente distantes.

Por un lado, la tioredoxina de *E. coli*, una enzima, que como ya hemos mencionado tiene actividad óxido-reductasa y puede reducir puentes disulfuro (catalizando la reacción $R_1-S-S-R_1 \rightarrow 2 R_1-SH$), por otro lado el dominio TRX de la chaperona *erp29*, que no tiene actividad óxido reductasa sino que interviene en el plegado de otras proteínas. Al comparar sus secuencias se observan grandes diferencias, entre ellas, los residuos del sitio activo de las tioredoxinas no están conservados en *erp29*. Al alinear las estructuras tridimensionales de ambas proteínas sin embargo, podemos ver que ambos plegados son superponibles. El alineamiento estructural fue realizado con el programa SwissPDBviewer. El alineamiento de estructuras primarias fue realizado con el programa ClustalX. (ver Figura 5-2)

Una pregunta muy importante que podemos hacernos es si dada una secuencia podemos, aunque más no sea, predecir qué estructura secundaria adoptaría. La respuesta es que, sí. Por lo general podemos.

En la actualidad existen numerosos métodos para hacerlo. Los más precisos utilizan simultáneamente varios tipos de información:

Cabe destacar que la conservación de un residuo particular puede volver más eficiente a una proteína desde un punto de vista funcional, pero simultáneamente puede desestabilizarla desde un punto de vista termodinámico-estructural

a. Puntual. Examina la tendencia de cada residuo en particular. De qué aminoácido se trata y cuál es la probabilidad de encontrarlo, formando parte de tal o cual elemento de estructura secundaria en las proteínas conocidas.

b. Local. Examina qué probabilidad hay de encontrarlo formando parte de tal o cual elemento dado su entorno inmediato, por ejemplo si está rodeado por residuos muy propensos a formar hélice o, por el contrario no lo está.

c. Evolutiva. Examinan en las secuencias homólogas disponibles en las bases de datos qué residuos ocupan esa posición y las vecinas.

Claramente, contar con una buena predicción de la estructura secundaria es un gran avance. Sin embargo,

¿Cómo interaccionan entre sí estos elementos de estructura secundaria, qué disposición adoptan en el espacio?

Si existe por lo menos una proteína homóloga con estructura tridimensional conocida podríamos montar nuestra secuencia sobre la estructura e intentar construir un modelo, aunque sea parcial, de la estructura de la proteína de interés.

Sin embargo tendremos un problema grave cuando no encontremos un homólogo que permitan construir un modelo confiable. En estos casos podemos recurrir a otro tipo de modelado. El modelado *ab initio* (desde el principio, aplicando primeros principios). Este tipo de modelado se basa en que gran parte de la información para la estructura 3D estaría localmente codificada. Esto quiere decir que cada pequeño fragmento de la proteína puede modelarse en base a bibliotecas de fragmentos existentes y, teniendo en cuenta predicciones de estructura secundaria, predicciones de accesibilidad a solvente, restricciones conformacionales, propias de las distintas cadenas laterales podríamos construir un modelo sin contar con la estructura de una proteína homóloga. Los resultados que se han obtenido son exitosos, sólo, parcialmente. Y se hace muy complicada la implementación de este tipo de metodología en forma rutinaria a pesar de que existen en la actualidad varios servidores funcionando. En definitiva todavía no es posible predecir la estructura de una proteína a partir de su secuencia si no contamos con un homólogo de estructura conocida.

Este tipo de estrategias suele utilizarse y se conocen en general con el nombre de modelado por homología.

Existen algunos problemas que complican aún más las cosas:

- a. Existen proteínas con secuencias muy diferentes (por debajo del límite de tección de homólogos) de aminoácidos con estructuras tridimensionales prácticamente superponibles.
- b. Existen proteínas con secuencias muy similares pero que poseen estructuras muy diferentes.
- c. Existen proteínas naturalmente desestructuradas.

Este último punto es sumamente importante porque, en la actualidad, se piensa que numerosas familias de proteínas podrían caer en este rango de la clasificación. La caracterización estructural de este tipo de macromoléculas es sumamente compleja, justamente, por la presencia del alto grado de desorden interno que poseen y las dificultades que esto trae aparejado en relación a los requerimientos que una proteína debe cumplir para poder hacer cristalografía de rayos X y NMR. En este caso estamos frente a un nuevo desafío:

¿Cómo estudiar proteínas desorganizadas?

Encontrar el código para el plegado parece ser una tarea complicada, sin embargo, la información requerida para que cada una de estas macromoléculas adopte una conformación tridimensional específica está contenida en la secuencia de aminoácidos.

La función biológica en realidad, depende no sólo de una proteína sino de una intrincada red de miles de proteínas que deben plegarse correctamente. Por lo general las proteínas adquieren la estructura correcta. Sin embargo existe una gran variedad de enfermedades que se caracterizan por la presencia de defectos de plegado; en una parte importante de los casos, estos desórdenes resultan de la acumulación de proteínas en forma de depósitos en los tejidos como consecuencia de la agregación proteica aparentemente irreversible.

Estos procesos son, en su conjunto, enormemente tóxicos para el organismo e inductores de procesos de muerte celular.

La conformación incorrectamente plegada, usualmente, contiene láminas β organizadas en un arreglo multimérico que tienden, naturalmente, a asociarse y estabilizarse mediante interacciones intermoleculares.

Existen mutaciones en los genes que codifican a estas proteínas que producen formas he-

Podríamos citar numerosos avances realizados por grupos de investigación distribuidos por todo el mundo. Se piensa que el código no es secuencial, sino que es espacial, y muchas pruebas indican que no toda la secuencia es esencial para el plegado. Encontrar el código de plegado no es un mero capricho.

Algunas de las enfermedades de tipo conformacional más conocidas y estudiadas a nivel molecular son las enfermedades de Alzheimer y Parkinson, la enfermedad de Huntington, las encefalopatías espongiformes transmisibles, la anemia hemolítica, la fibrosis quística y la diabetes de tipo II.

Aunque cada una de estas enfermedades es causada por defectos conformacionales asociados una proteína distinta con una secuencia de aminoácidos particular, todas estas proteínas tiene un aspecto común: puede adoptar por lo menos dos conformaciones.

reditarias y muy graves de estas enfermedades. Por otro lado, en el caso particular de las encefalopatías espongiformes transmisibles pudo demostrarse que la enfermedad puede transmitirse a través de mecanismos de cambio conformacional y agregación que involucran agentes proteicos infecciosos capaces de modular la conformación de proteínas propias del organismo que es infectado.

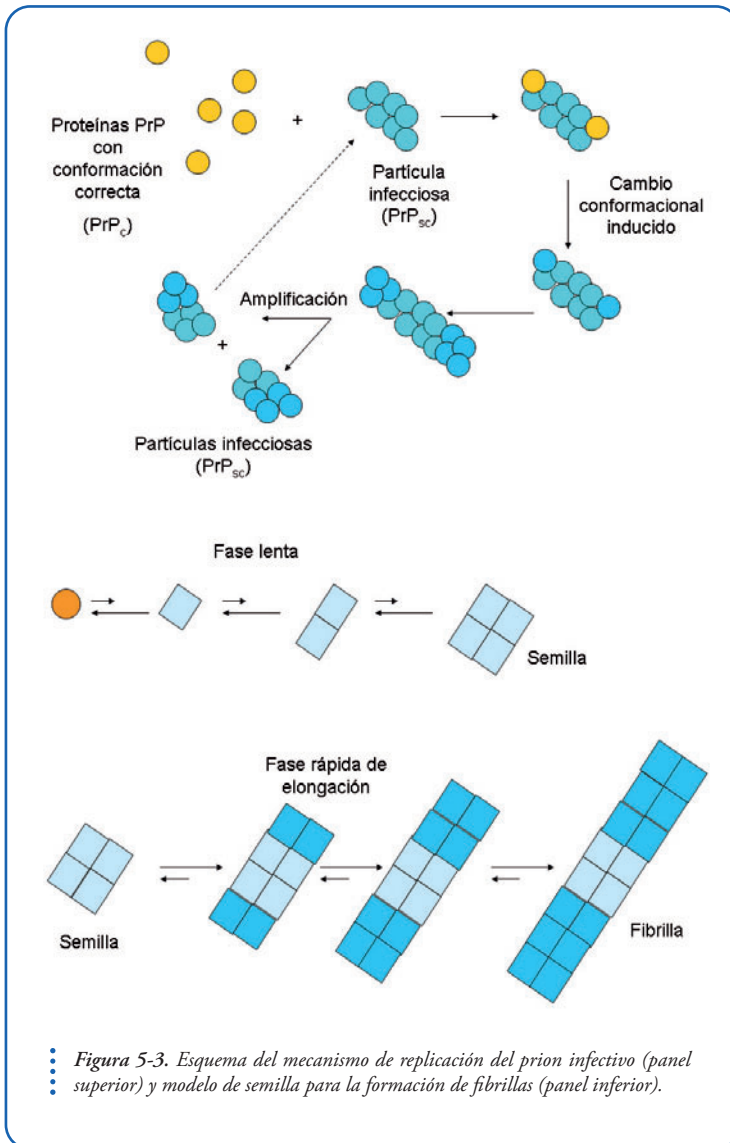
Hasta hace unos pocos años era impensado que una proteína pudiera estar involucrada como único agente infeccioso en procesos infecciosos. Ahora comenzamos a conocer partes de un mecanismo, sumamente, intrigante en el que, efectivamente, el agente infeccioso es una proteína incorrectamente plegada. Las encefalopatías espongiformes afectan tanto a animales como a humanos y se caracterizan por producir muerte neuronal (por lo que se las llaman enfermedades neurodegenerativas) y acumulación de una proteína llamada prion (PrP) incorrectamente plegada llamada PrP_{sc} en el sistema nervioso central.

Aunque el mecanismo preciso a nivel molecular de conversión de PrP_c a PrP_{sc}, la conformación correcta en la conformación PrP_{sc} no se conoce, las evidencias indican que PrP_{sc} es un oligómero que actúa como semilla para unir a la PrP_c y catalizar el cambio conformacional a la forma incorrectamente plegada mediante la incorporación al multímero creciente (**Figura 5-3**).

En esta breve lista se enumeran las evidencias que avalan la hipótesis del cambio conformacional de la proteína prion.

En la actualidad se piensa que muchas proteínas tienen la capacidad de funcionar como priones y que, como se mencionó, las formas infectivas difieren de las no infectivas, únicamente, en su estructura tridimensional. Sin embargo hasta, recientemente, no disponíamos de estructuras del estado de fibra que, como también se mencionó, muy probablemente sea infeccioso. Por técnicas de resonancia magnética nuclear (NMR) a fines del año 2008 se estudió un dominio de la proteína HET-S que posee propiedades de prion y se obtuvieron las primeras estructuras de este tipo de conformaciones (Christian Wasmer y colaboradores **Science 14 March 2008: Vol. 319. no. 5869, pp. 1523 – 1526**). Los resultados de estos experimentos muestran la existencia de un alto contenido de estructura β y a las subunidades dispuestas en forma de solenoide. En este caso el modelo experimental permite inferir cómo interactúan las subunidades entre sí para formar estructuras con forma de fibrilla (**ver Figura 5-4**).

Estas enfermedades, en su origen raras, han ganado popularidad con la emergencia en Europa (Inglaterra) de la nueva variante de la enfermedad Creutzfeldt–Jacob causada por el consumo de carne infectada proveniente de animales con encefalopatía espongiforme bovina. La naturaleza proteica del agente infeccioso y el mecanismo de propagación fueron descubiertos por el Profesor de Neurología y Bioquímica de la Universidad de California, San Francisco Stanley B. Prusiner; estos descubrimientos le valieron en 1997 el Premio Nobel Medicina. Actualmente, todas las evidencias sugieren el extraordinario concepto de que la proteína incorrectamente plegada PrP_{sc} es el único componente necesario productor de la infección.



La formación de las semillas es lenta porque requeriría un cambio conformacional relativamente poco probable. Sin embargo, una vez generadas, estas semillas actúan como núcleo para una posterior y rápida conversión a la forma PrP_{Sc} de las moléculas. (ver Figura 5-3)

Estas evidencias han sido citadas en una excelente revisión realizada recientemente por Claudio Soto y colaboradores (TRENDS in Biochemical Sciences Vol.31 N° 3 marzo de 2006):

(I) El agente infeccioso no es afectado por procedimientos que normalmente destruyen a los ácidos nucleicos

Alper, T. et al. (1967) Does the agent of scrapie replicate without nucleic acid? Nature 214, 764–766

(II) El material infeccioso es muy pequeño para ser un microorganismo convencional.

Alper, T. et al. (1966) The exceptionally small size of the scrapie agent. Biochem. Biophys. Res. Commun. 22, 278–284

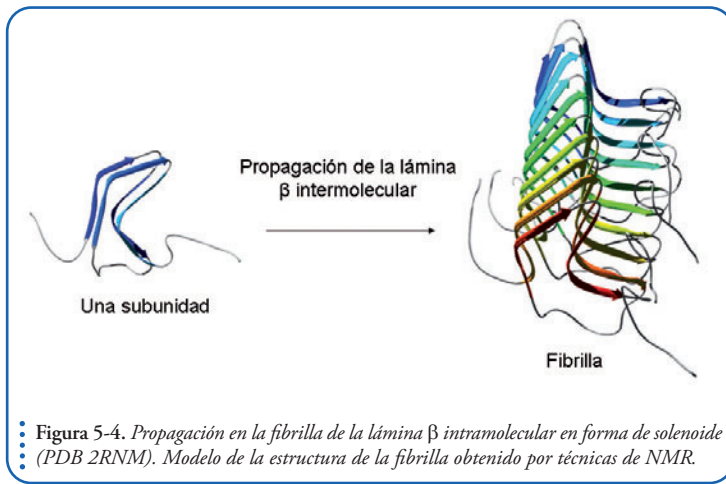
Silveira, J.R. et al. (2005) The most infectious prion protein particles. Nature 437, 257–261

(III) No se ha podido asociar partículas virales con el agente infeccioso

Safar, J.G. et al. (2005) Search for a prion-specific nucleic Acid. J. Virol. 79, 10796–10806

(IV) Preparaciones de proteína PrP_{Sc} altamente purificada transmiten la enfermedad

Prusiner, S.B. et al. (1984) Purification and structural studies of a major scrapie prion protein. Cell 38, 127–134



Desgraciadamente, en la actualidad, no podemos decir que contamos con las soluciones y curas para estas y tantas otras enfermedades, sin embargo, el estudio estructural muy probablemente, permitirá comprender las bases moleculares de los mecanismos de cambio conformacional y tal vez, con suerte la elaboración de soluciones.

(V) La infectividad es proporcional a la concentración de la proteína PrP^{Sc}.

Gabizon, R. et al. (1988) Immunoaffinity purification and neutralization of scrapie prion infectivity. *Proc. Natl. Acad. Sci. U. S. A.* 85, 6617–6621

(VI) Procedimientos de desnaturación agresiva de PrP^{Sc} reducen o eliminan la infectividad.

Prusiner, S.B. et al. (1993) Attempts to restore scrapie prion infectivity after exposure to protein denaturants. *Proc. Natl. Acad. Sci. U. S. A.* 90, 2793–2797

(VII) Los ratones modificados genéticamente de tal forma que se les elimina el gen propio de PrP^C no son susceptibles a la infección.

Bueler, H. et al. (1993) Mice devoid of PrP are resistant to scrapie. *Cell* 73, 1339–1347

(VIII) En todos los casos en los que la enfermedad es hereditaria existen mutaciones del gen PrP

Prusiner, S.B. (1998) Prions. *Proc. Natl. Acad. Sci. U. S. A.* 95, 13363–13383

(IX) PrP^{Sc} puede inducir el cambio conformacional de PrP^C in-vitro en forma autocatalítica.

Saborio, G.P. et al. (2001) Sensitive detection of pathological prion protein by cyclic amplification of protein misfolding. *Nature* 411, 810–813

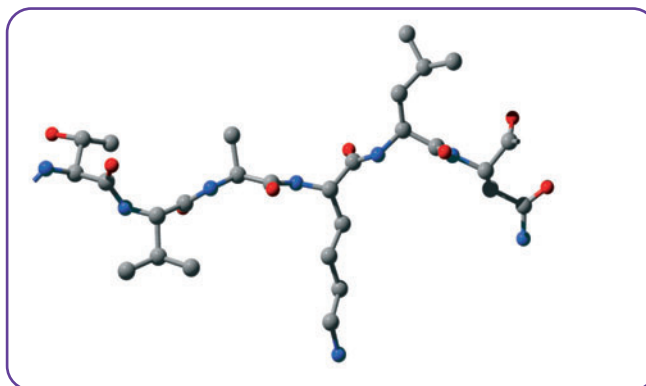
Guía de problemas

Problema P1. Identificar en la Figura la secuencia de aminoácidos del péptido. Encerrar con círculos punteados los Ca. Tener en cuenta que no se representan los átomos de hidrógeno.

Problema P2. Obtener a través de la web el archivo del modelo experimental de la estructura de la frataxina humana, código PDB: 1EKG.

Hay alguna dirección para recomendar?

Problema P3. Abrir el archivo 1EKG.pdb con un procesador de textos (por ejemplo Word). ¿De dónde sacan el archivo? Reconocer en este extenso archivo las líneas en las que se almacenan las coordenadas espaciales de los átomos del modelo (consultar Capítulo I). Identificar en el fragmento del archivo que se transcribe abajo cuáles átomos pertenecen a la cadena principal de la proteína y cuáles corresponden a las cadenas laterales de los aminoácidos.

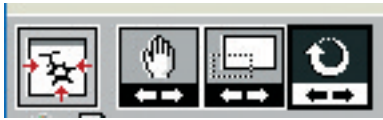


La dirección del banco de estructuras de proteínas es: <http://www.rcsb.org/pdb/home/home.d>

ATOM	1	N	LEU	A	90	1.020	13.240	5.227	1.00	37.83	N
ATOM	2	CA	LEU	A	90	-0.329	12.612	5.341	1.00	37.74	C
ATOM	3	C	LEU	A	90	-1.386	13.607	4.894	1.00	35.93	C
ATOM	4	O	LEU	A	90	-1.081	14.771	4.628	1.00	36.75	O
ATOM	5	CB	LEU	A	90	-0.633	12.224	6.794	1.00	39.88	C
ATOM	6	CG	LEU	A	90	0.491	11.857	7.765	1.00	40.72	C
ATOM	7	CD ₁	LEU	A	90	1.365	10.769	7.161	1.00	42.11	C
ATOM	8	CD ₂	LEU	A	90	1.305	13.102	8.082	1.00	41.63	C
ATOM	9	N	ASP	A	91	-2.630	13.147	4.809	1.00	33.80	N
ATOM	10	CA	ASP	A	91	-3.725	14.027	4.436	1.00	31.98	C
ATOM	11	C	ASP	A	91	-4.099	14.771	5.713	1.00	30.82	C
ATOM	12	O	ASP	A	91	-3.594	14.442	6.790	1.00	29.10	O
ATOM	13	CB	ASP	A	91	-4.913	13.227	3.895	1.00	33.03	C
ATOM	14	CG	ASP	A	91	-5.405	12.176	4.864	1.00	35.00	C

ATOM	15	OD ₁	ASP	A	91	-6.415	11.517	4.551	1.00	38.57	O
ATOM	16	OD ₂	ASP	A	91	-4.787	12.001	5.931	1.00	37.11	O
ATOM	17	N	GLU	A	92	-4.977	15.763	5.614	1.00	29.58	N
ATOM	18	CA	GLU	A	92	-5.332	16.542	6.793	1.00	29.88	C
ATOM	19	C	GLU	A	92	-6.086	15.782	7.880	1.00	27.15	C
ATOM	20	O	GLU	A	92	-5.938	16.087	9.059	1.00	26.33	O
ATOM	21	CB	GLU	A	92	-6.112	17.802	6.393	1.00	32.28	C
ATOM	22	CG	GLU	A	92	-7.610	17.640	6.284	1.00	37.12	C
ATOM	23	CD	GLU	A	92	-8.320	18.985	6.199	1.00	41.17	C
ATOM	24	OE ₁	GLU	A	92	-9.566	18.996	6.089	1.00	43.24	O
ATOM	25	OE ₂	GLU	A	92	-7.629	20.031	6.246	1.00	39.40	O
ATOM	26	N	THR	A	93	-6.887	14.796	7.493	1.00	26.15	N
ATOM	27	CA	THR	A	93	-7.644	14.011	8.464	1.00	25.69	C
ATOM	28	C	THR	A	93	-6.707	13.195	9.345	1.00	23.63	C
ATOM	29	O	THR	A	93	-6.869	13.141	10.570	1.00	22.64	O
ATOM	30	CB	THR	A	93	-8.619	13.047	7.763	1.00	28.32	C
ATOM	31	OG ₁	THR	A	93	-9.614	13.807	7.067	1.00	31.76	O
ATOM	32	CG ₂	THR	A	93	-9.297	12.131	8.780	1.00	30.09	C
ATOM	33	N	THR	A	94	-5.723	12.557	8.721	1.00	21.85	N
ATOM	34	CA	THR	A	94	-4.768	11.753	9.470	1.00	20.80	C
ATOM	35	C	THR	A	94	-3.926	12.674	10.348	1.00	19.41	C
ATOM	36	O	THR	A	94	-3.700	12.386	11.522	1.00	17.85	O
ATOM	37	CB	THR	A	94	-3.850	10.959	8.529	1.00	23.10	C
ATOM	38	OG ₁	THR	A	94	-4.657	10.179	7.635	1.00	22.21	O
ATOM	39	CG ₂	THR	A	94	-2.941	10.028	9.328	1.00	19.76	C
ATOM	40	N	TYR	A	95	-3.468	13.783	9.776	1.00	17.80	N
ATOM	41	CA	TYR	A	95	-2.670	14.740	10.535	1.00	17.60	C
ATOM	42	C	TYR	A	95	-3.415	15.168	11.800	1.00	18.28	C
ATOM	43	O	TYR	A	95	-2.856	15.165	12.901	1.00	16.97	O
ATOM	44	CB	TYR	A	95	-2.372	15.993	9.709	1.00	16.23	C
ATOM	45	CG	TYR	A	95	-1.974	17.152	10.592	1.00	17.96	C
ATOM	46	CD ₁	TYR	A	95	-0.696	17.216	11.155	1.00	16.22	C
ATOM	47	CD ₂	TYR	A	95	-2.913	18.104	10.984	1.00	16.36	C
ATOM	48	CE ₁	TYR	A	95	-0.373	18.190	12.096	1.00	17.51	C
ATOM	49	CE ₂	TYR	A	95	-2.599	19.083	11.925	1.00	17.43	C
ATOM	50	CZ	TYR	A	95	-1.330	19.116	12.480	1.00	16.66	C
ATOM	51	OH	TYR	A	95	-1.028	20.042	13.452	1.00	17.76	O

Problema P4. Obtener a través de la web en <http://spdbv.vital-it.ch/> el programa SwissPDBviewer. Ejecutar el programa. Abrir con este programa el archivo 1ekg.pdb. Verificar el funcionamiento de las herramientas generales. Sacar conclusiones. Íconos de herramientas generales:



Problema P5. Verificar el funcionamiento del panel de control y las posibilidades de representación de la cadena polipeptídica que brinda este panel: modelo de cintas (*ribbons*), visualización de átomos de cadena principal (*backbone*) y de cadenas laterales (*side-chains*).

Generar el modelo de cintas. Colorear en rojo los átomos que corresponden a los primeros 24 residuos del modelo 1ekg: Leu90 a Ala114. Colorear la sección del modelo de cintas que va desde Leu90 a la Ala114, que corresponde a la primera hélice α .

Problema P6. Determinar aritméticamente la distancia entre carbonos α de los residuos Leu A 90 y Asp A 91 de la frataxina humana utilizando las coordenadas espaciales x,y,z . Rehacer el cálculo pero ahora con la ayuda del programa de visualización de estructuras SwissPDBviewer.

ATOM	2	CA	LEU	A	90	-0.329	12.612	5.341	1.00	37.74	C
ATOM	10	CA	ASP	A	91	-3.725	14.027	4.436	1.00	31.98	C

Problema P7. Determinar cuáles son los residuos vecinos del Trp173 de la frataxina humana (1ekg.pdb) dado un radio $r=4.5\text{\AA}$.

Problema P8. Construir el gráfico de Ramachandran con el programa *SwissPDBViewer* para 1ekg.

Problema P9. Construir un modelo extendido de la cadena polipeptídica en el que la exposición de los aminoácidos al solvente sea máxima.

Problema P10. Definir si los residuos Ser158 y Pro159 del modelo de la frataxina humana 1ekg se encuentran unidos a través de un enlace peptídico en configuración *cis* o *trans*.

Problema P11. Calcular la superficie accesible al solvente del modelo 1ekg con el programa SwissPDBviewer.

Problema P12. En la **Figura P12** se muestra un experimento de desnaturalización en el equilibrio de la tiorredoxina de *E. coli* (variante salvaje) y de una mutante, la variante L107A. En la **Tabla 1** se muestran los parámetros termodinámicos correspondientes a un ajuste a dos estados ($N \leftrightarrow U$).

Tabla 1. Parámetros termodinámicos.

	ΔG_{NU} (kcal Mol ⁻¹)	m_{NU} (kcal mol ⁻¹ M ⁻¹)	S_N (Int. U.A.) ¹	S_U (Int. U.A.)	C_m (M)
TRX _{salvaje}	7,9	2,95	47,32	169,92	?
TRXL _{107A}	6,2	3,28	53,54	175,22	?

¹Int.U.A es intensidad de fluorescencia en unidades arbitrarias.

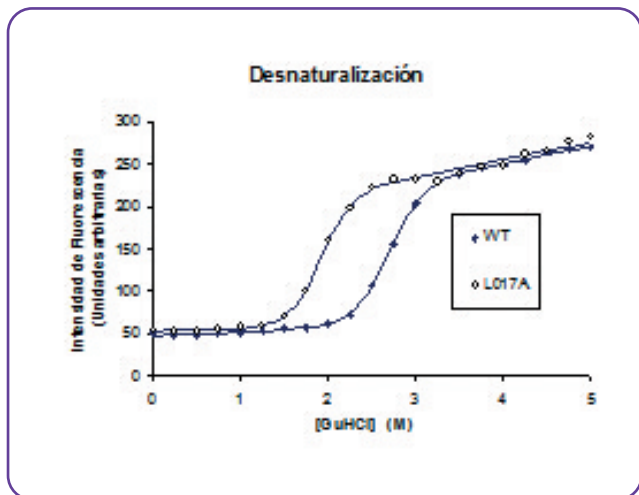


Figura P12. Desnaturalización de la tiorredoxina de *E. coli*.

- Calcular la C_m (concentración media de desnaturalizante) para ambos procesos. ¿Qué valor toma el ΔG_{NU} en esta concentración de desnaturalizante?
- ¿Qué variante es más estable?
- ¿Cómo son las constantes de velocidad del proceso de repliegado y de desplegado en la C_m ?

Problema P13

- ¿Cuál es la absorbancia a 280 nm de una solución 35 μM de triptofano?
- ¿y de una solución del péptido AKSTEF? ¿y del péptido AKSTEYY?
- Si tuviera los péptidos en su forma pura y en dos tubos sin rótulo, podría identificarlos por sus espectros UV? ¿Cómo? Dato: se usa una cubeta de 1 cm de paso óptico.

Problema P14. Se preparó una muestra de proteína para hacer un experimento de espectroscopía de fluorescencia de triptofano. La absorbancia de la solución a 280 nm es de 0.1.

- Podría determinar la concentración en:
 - M
 - mg/mL
- Calcular el número de moléculas presentes en la cubeta de medición.

Datos:

$$24750 \text{ M}^{-1} \text{ cm}^{-1} = \epsilon_{280\text{nm}}$$

1 mL = Volumen de la cubeta de medición

1 cm = Paso óptico de cubeta usada en el espectrofotómetro para medir la absorbancia

29500 Da = Peso molecular de la proteína utilizada.

Problema P15. Explique por qué el triptofano libre en solución posee un espectro de fluorescencia con un $\lambda_{MAX} = 356 \text{ nm}$, mientras que en el contexto proteico del estado nativo, el λ_{MAX} puede ser menor, por ejemplo 340 nm. En un diagrama de Jablonski represente este proceso y agregue la escala temporal de los procesos de absorción y fluorescencia.

Problema P16. ¿Qué tipo de información podemos obtener a partir de resultados de experimentos de dicroísmo circular?

Problema P17. Calcular la masa de la proteína de unión a ácidos grasos utilizando como datos las relaciones m/z del ión más abundante y del inmediatamente siguiente a su izquierda (Figura 4-48).

$$\frac{m}{z} = 1000,543$$

$$\frac{m}{z} = 938,082$$

Problema 18. Se quiere analizar la estructura cuaternaria de una proteína cuya masa es de 10063.5 Da.

Con este fin se realiza un experimento de SEC-FPLC y se usa una columna Superdex S-200. Se muestra un resumen del perfil cromatográfico correspondiente a los marcadores de tamaño molecular en la tabla. El volumen de elución del complejo macromolecular es de 13.76 ml. a. ¿Si la proteína fuera monomérica y globular, que R_s tendría?, en ese caso, b. ¿Cuál debería ser el volumen de elución asociado? c. ¿Qué puede decir, sin embargo de la estructura cuaternaria de este complejo?

PM	Log R_s	R_s	Tiempo de retención	Volumen de elución
670000	1.89	78.67	17.84	8.92
150000	1.65	45.28	24.05	12.02
44000	1.45	28.80	28.99	14.49
17000	1.30	20.27	33.56	16.78
11692	1.24	17.66	35.31	17.65
1350	0.90	7.96	40.6	20.30

Problema P19. Una proteína recientemente clonada posee 3 fenilalaninas pero ningún triptófano y ninguna tirosina. En el laboratorio se prepararon 3 mutantes en cada una de ellas una de las fenilalaninas se reemplaza por triptófano F34W (—), F80W (••••) y F101W (---). Prediga a partir de los resultados obtenidos por apagamiento de la fluorescencia de triptófano por acrilamida cuál de las Phe es la posee menor accesibilidad al solvente.

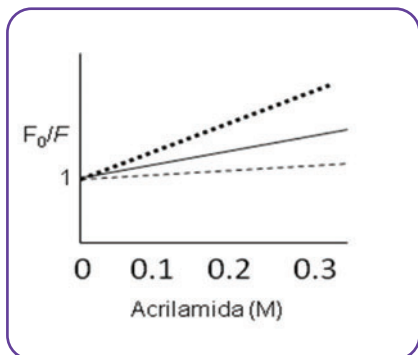


Figura P19 Apagamiento de la fluorescencia de Trp por acrilamida. F34W (—), F80W (••••) y F101W (---). F0 es la Fluorescencia en ausencia de acrilamida F es la fluorescencia en presencia de la concentración molar de acrilamida indicada.

Problema P20. Se dispone de dos proteínas. Indicar cuál de los dos espectros de fluorescencia de triptófano corresponde al estado nativo y cuál al estado desplegado en cada caso. Dar una explicación para los resultados observados.

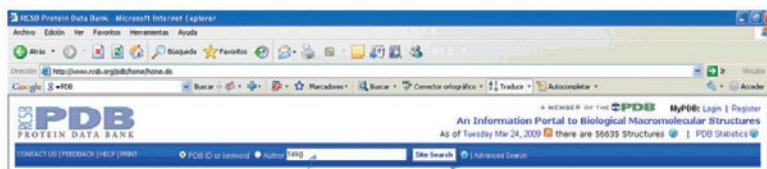
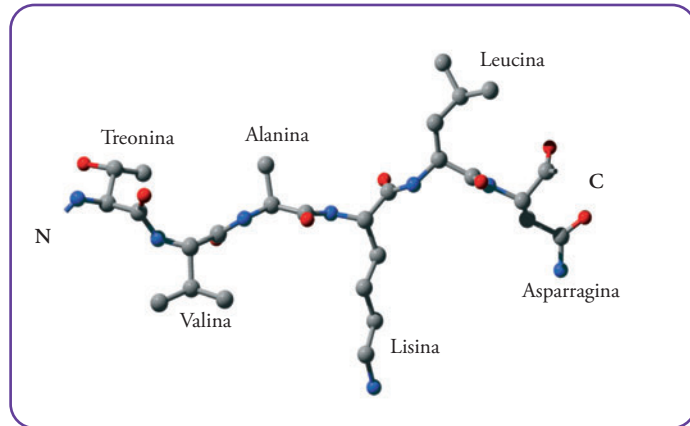
Respuestas de problemas

Respuesta P1.

Se trata del hexapéptido Thr-Val-Ala-Lys-Leu-Asn. Con un círculo se resaltan los carbonos alfa y se indican los extremos N- y C- terminales con flechas.

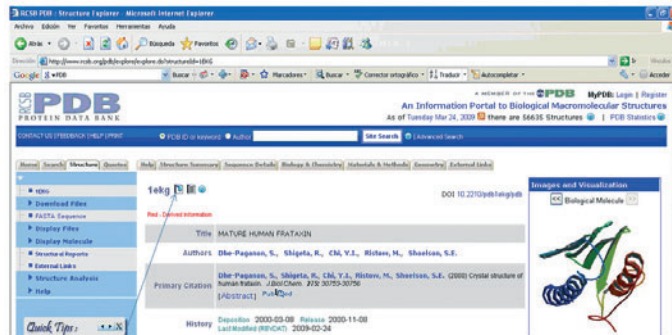
Respuesta P2.

Se podrá acceder a la PDB y luego se deberá guardar el archivo 1ekg como archivo de texto pero con la terminación .pdb siguiendo las instrucciones de la Figura P2-1, paneles superior e inferior.



(i) Introducir código de identificación PDB: 1ekg

(ii) Iniciar la búsqueda de la estructura haciendo un clic aquí



(iii) Haciendo un doble clic aquí se inicia la bajada del archivo 1ekg.pdb

Respuesta P3.

Los átomos correspondientes a la cadena principal se resaltan en amarillo.

ATOM	1	N	LEU	A	90	1.020	13.240	5.227	1.00	37.83	N
ATOM	2	CA	LEU	A	90	-0.329	12.612	5.341	1.00	37.74	C
ATOM	3	C	LEU	A	90	-1.386	13.607	4.894	1.00	35.93	C
ATOM	4	O	LEU	A	90	-1.081	14.771	4.628	1.00	36.75	O
ATOM	5	CB	LEU	A	90	-0.633	12.224	6.794	1.00	39.88	C
ATOM	6	CG	LEU	A	90	0.491	11.857	7.765	1.00	40.72	C
ATOM	7	CD ₁	LEU	A	90	1.365	10.769	7.161	1.00	42.11	C
ATOM	8	CD ₂	LEU	A	90	1.305	13.102	8.082	1.00	41.63	C
ATOM	9	N	ASP		91	-2.630	13.147	4.809	1.00	33.80	N
ATOM	10	CA	ASP		91	-3.725	14.027	4.436	1.00	31.98	C
ATOM	11	C	ASP		91	-4.099	14.771	5.713	1.00	30.82	C
ATOM	12	O	ASP		91	-3.594	14.442	6.790	1.00	29.10	O
ATOM	13	CB	ASP	A	91	-4.913	13.227	3.895	1.00	33.03	C
ATOM	14	CG	ASP	A	91	-5.405	12.176	4.864	1.00	35.00	C
ATOM	15	OD ₁	ASP	A	91	-6.415	11.517	4.551	1.00	38.57	O
ATOM	16	OD ₂	ASP	A	91	-4.787	12.001	5.931	1.00	37.11	O
ATOM	17	N	GLU		92	-4.977	15.763	5.614	1.00	29.58	N
ATOM	18	CA	GLU		92	-5.332	16.542	6.793	1.00	29.88	C
ATOM	19	C	GLU		92	-6.086	15.782	7.880	1.00	27.15	C
ATOM	20	O	GLU		92	-5.938	16.087	9.059	1.00	26.33	O
ATOM	21	CB	GLU	A	92	-6.112	17.802	6.393	1.00	32.28	C
ATOM	22	CG	GLU	A	92	-7.610	17.640	6.284	1.00	37.12	C
ATOM	23	CD	GLU	A	92	-8.320	18.985	6.199	1.00	41.17	C
ATOM	24	OE ₁	GLU	A	92	-9.566	18.996	6.089	1.00	43.24	O
ATOM	25	OE ₂	GLU	A	92	-7.629	20.031	6.246	1.00	39.40	O
ATOM	26	N	THR		93	-6.887	14.796	7.493	1.00	26.15	N
ATOM	27	CA	THR		93	-7.644	14.011	8.464	1.00	25.69	C
ATOM	28	C	THR		93	-6.707	13.195	9.345	1.00	23.63	C
ATOM	29	O	THR		93	-6.869	13.141	10.570	1.00	22.64	O
ATOM	30	CB	THR	A	93	-8.619	13.047	7.763	1.00	28.32	C
ATOM	31	OG ₁	THR	A	93	-9.614	13.807	7.067	1.00	31.76	O
ATOM	32	CG ₂	THR	A	93	-9.297	12.131	8.780	1.00	30.09	C
ATOM	33	N	THR		94	-5.723	12.557	8.721	1.00	21.85	N
ATOM	34	CA	THR		94	-4.768	11.753	9.470	1.00	20.80	C
ATOM	35	C	THR		94	-3.926	12.674	10.348	1.00	19.41	C
ATOM	36	O	THR		94	-3.700	12.386	11.522	1.00	17.85	O
ATOM	37	CB	THR	A	94	-3.850	10.959	8.529	1.00	23.10	C
ATOM	38	OG ₁	THR	A	94	-4.657	10.179	7.635	1.00	22.21	O
ATOM	39	CG ₂	THR	A	94	-2.941	10.028	9.328	1.00	19.76	C
ATOM	40	N	TYR	A	95	-3.468	13.783	9.776	1.00	17.80	N

ATOM	41	CA	TYR	A	95	-2.670	14.740	10.535	1.00	17.60	C
ATOM	42	C	TYR	A	95	-3.415	15.168	11.800	1.00	18.28	C
ATOM	43	O	TYR	A	95	-2.856	15.165	12.901	1.00	16.97	O
ATOM	44	CB	TYR	A	95	-2.372	15.993	9.709	1.00	16.23	C
ATOM	45	CG	TYR	A	95	-1.974	17.152	10.592	1.00	17.96	C
ATOM	46	CD ₁	TYR	A	95	-0.696	17.216	11.155	1.00	16.22	C
ATOM	47	CD ₂	TYR	A	95	-2.913	18.104	10.984	1.00	16.36	C
ATOM	48	CE ₁	TYR	A	95	-0.373	18.190	12.096	1.00	17.51	C
ATOM	49	CE ₂	TYR	A	95	-2.599	19.083	11.925	1.00	17.43	C
ATOM	50	CZ	TYR	A	95	-1.330	19.116	12.480	1.00	16.66	C
ATOM	51	OH	TYR	A	95	-1.028	20.042	13.452	1.00	17.76	O

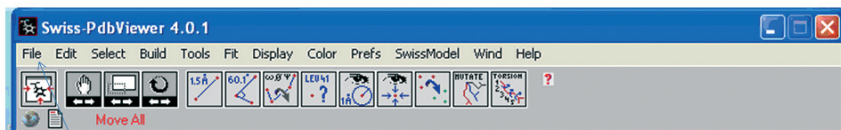
Respuesta P4.

Para ejecutar el programa hacer un doble clic en el ícono

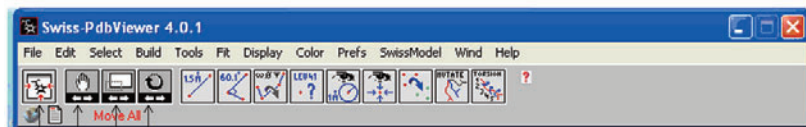


spdbv

Para abrir el archivo, 1ekg.pdb



(i) Hacer un clic
en File, *open PDB
file*



Girar
Ampliar
Desplazar horizontalmente
Centrar

Herramientas
generales

Respuesta P5.

El panel de control

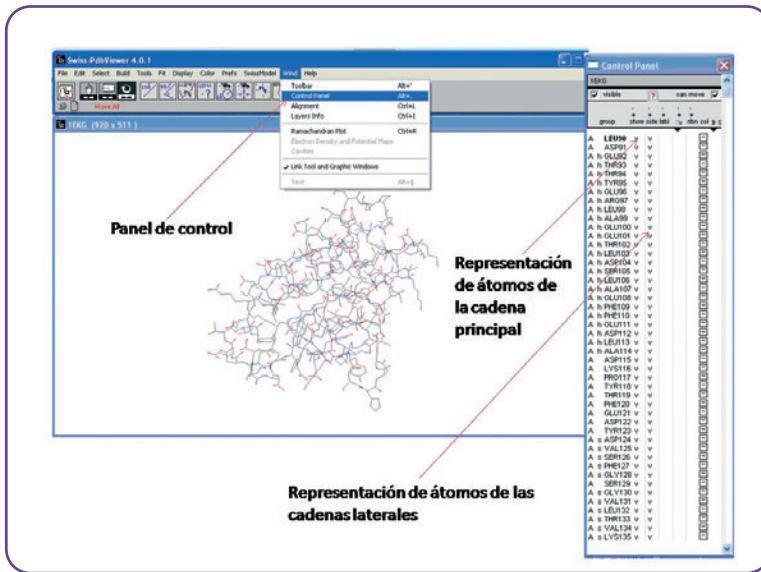
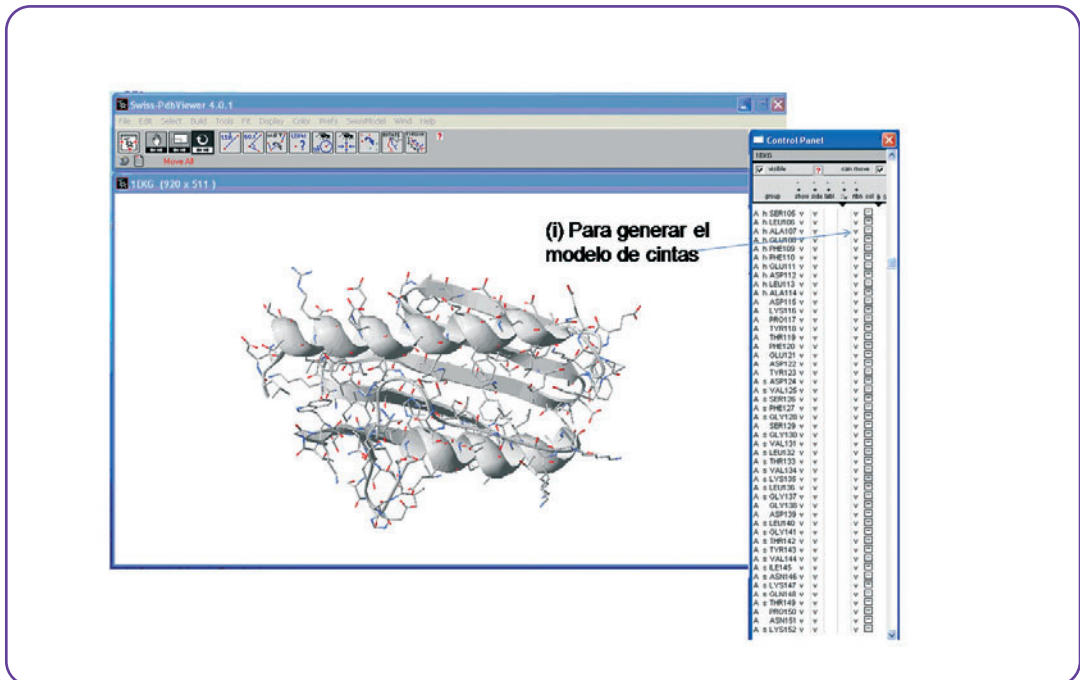


Figura P5-1. ¿Cómo visualizar el panel de control y qué columnas de este permiten representar a los átomos?

170



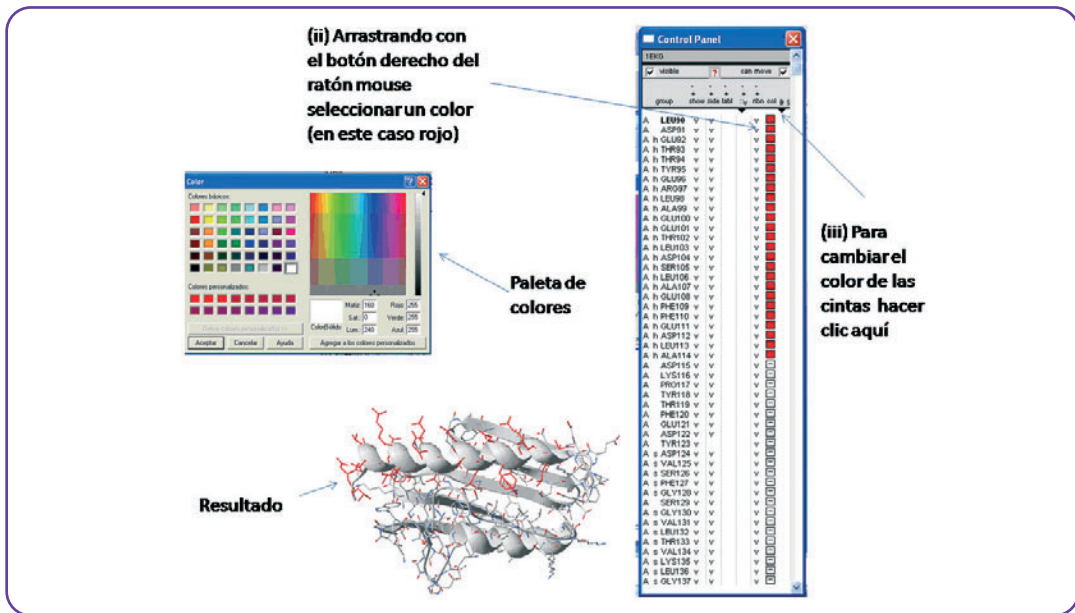


Figura P5-2.

Se muestra qué columna del panel permite representar el modelo en forma de cintas y cómo colorear determinados átomos del modelo, en este caso los que corresponden a la hélice $\alpha 1$ de la frataxina humana (1ekg.pdb).

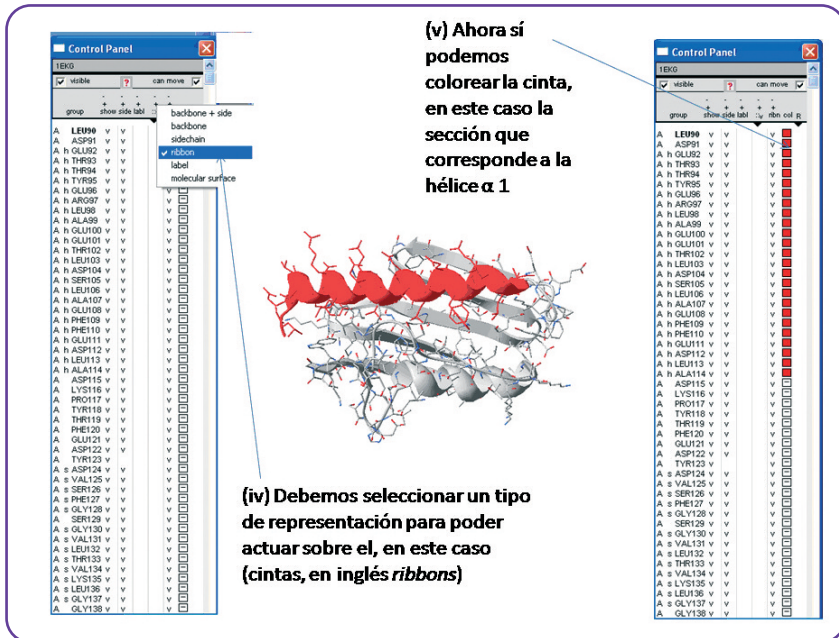


Figura P5-3. Cómo colorear una sección de cinta. En rojo se muestra la sección correspondiente a los primeros 24 residuos, hélice $\alpha 1$.

Respuesta P6.

Una pregunta muy común es *qué distancia separa a dos átomos en nuestro modelo*: por ejemplo distancia entre dos carbonos α (CA y CA). En las dos líneas de abajo, repetimos las coordenadas espaciales del carbono

alfa de la leucina 90 y del carbono alfa de ácido aspártico 91 de 1EKG. Las columnas 7,8 y 9 corresponden en un archivo PDB como hemos visto en el capítulo 1 a las coordenadas espaciales x, y, z

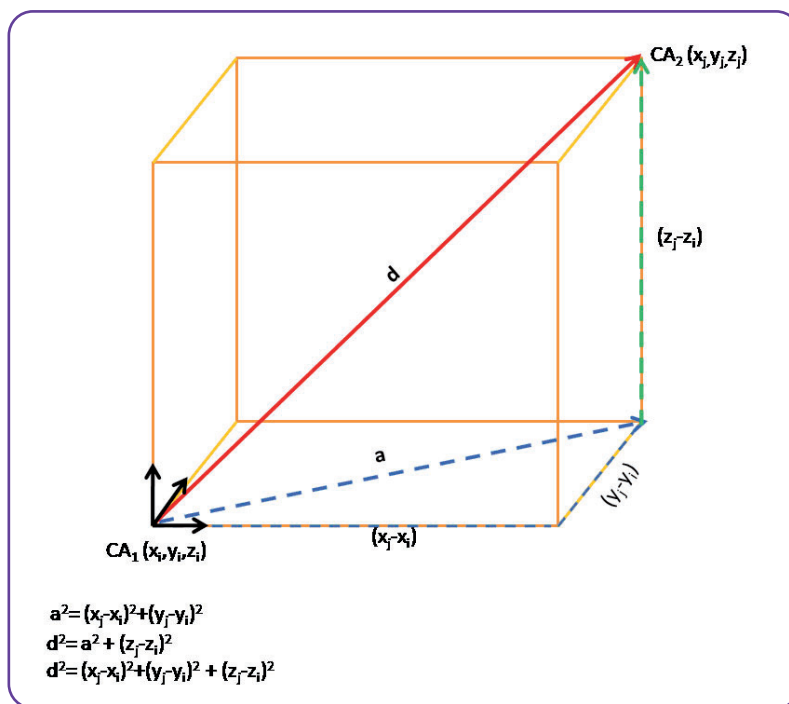
Aplicando el teorema de Pitágoras pero en un espacio tridimensional podemos muy rápidamente calcular esta distancia.

						x	y	z			
ATOM	2	CA	LEU	A	90	-0.329	12.612	5.341	1.00	37.74	C
ATOM	10	CA	ASP	A	91	-3.725	14.027	4.436	1.00	31.98	C

$$d^2 = (x_2 - x_{10})^2 + (y_2 - y_{10})^2 + (z_2 - z_{10})^2$$

$$d^2 = (-0.329 + 3.725)^2 + (12.612 - 14.027)^2 + (5.341 - 4.436)^2$$

$$d = 3.79 \text{ \AA}$$



En la Figura P6-1 se muestra una representación de la operación general para el cálculo de la distancia entre dos átomos: CA_j y CA_i.

Esta distancia puede determinarse muy fácilmente utilizando programas de visualización como el *SPDBviewer* (Figura P6-2). En este caso podemos abrir el archivo PDB de la frataxina desde la barra de herramientas del *SPDBviewer* (en *File, open PDB file*). Una vez abierto el archivo, podemos usar el *Control Panel* (panel de control) para visualizar únicamente los residuos que nos interesen y elegir el sistema de representación de átomos que nos convenga.

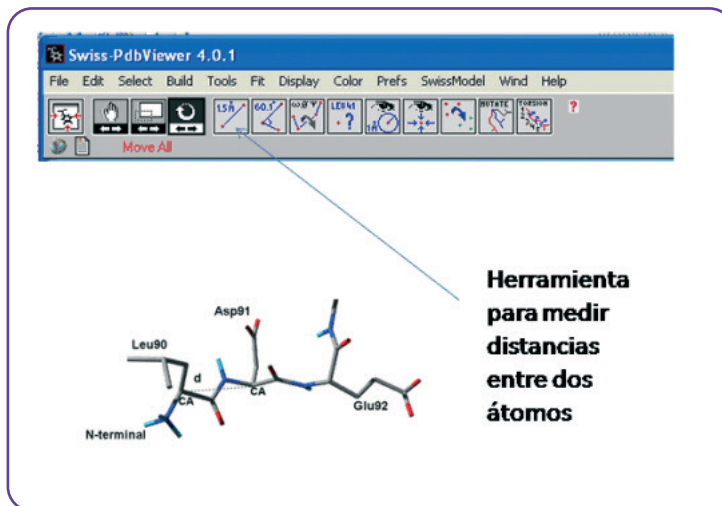
En la barra de herramientas principal podemos seleccionar la de medir distancias (cuando la seleccionemos se pintará de negro y aparecerá una leyenda: *Pick one atom*, deberemos entonces seleccionar de la imagen un átomo haciendo un clic sobre este (en nuestro ejemplo CA de Leu90) y después haciendo un segundo clic sobre el segundo CA de Asp91 (en rojo podrá leerse *Pick 2nd atom*) seleccionamos el segundo átomo que necesitamos para realizar la medición.

Figura P6-2.

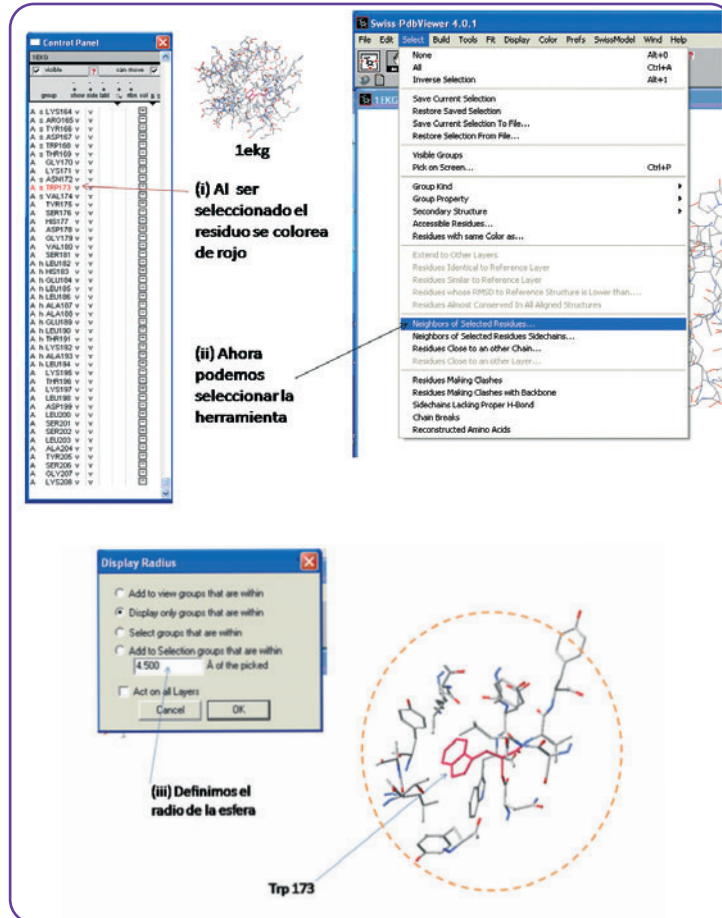
Determinación de la distancia entre dos átomos usando la barra de herramientas del programa *SwissPDBviewer*.

Respuesta P7.

Concretamente los residuos vecinos del residuo (i), son aquellos residuos que se encuentran completa o parcialmente dentro de una esfera de radio r centrada en el residuo (i). Si r es muy grande (por ejemplo 30Å) y la proteína es relativamente pequeña (100 aminoácidos) todos o la mayor parte de los residuos de la proteína quedarán incluidos en la esfera. En cambio si r es pequeño, por ejemplo 4 ó 5Å, la esfera incluirá únicamente a los vecinos inmediatos. Con el programa *SwissPdbViewer* resulta sencillo encontrar a los residuos vecinos dado un radio r (Figura P7-1 A y B). Para esto se selecciona un residuo haciendo un clic sobre el panel control (*control panel*), en este ejemplo seleccionamos al triptofano 173 de la frataxina, (Trp 173). En select, se busca el comando *neighbors of selected* aminoacids, (en castellano quiere decir vecinos del amino ácido seleccionado). Luego para que muestre únicamente los residuos que están dentro de la esfera (en este ejemplo de



Herramienta para medir distancias entre dos átomos



4,5 Å de radio) se utiliza el comando *Display only groups that are within*).

Figura P7- 1.

Caracterización de un entorno usando el programa *SwissPdbViewer*. En el panel A se muestra la

selección del residuo TRP 173 (se lo señala con una flecha roja). En el panel B se define el radio de la esfera, 4,5Å con su centro en el residuo 173. Entre los residuos vecinos podemos observar a la Leu98, Ala99, Thr102, Lys152 e Ile154. ¿Podrías identificarlos?

Para identificar un residuo en una estructura proteica podemos utilizar una de las herramientas del SwissPDBviewer.

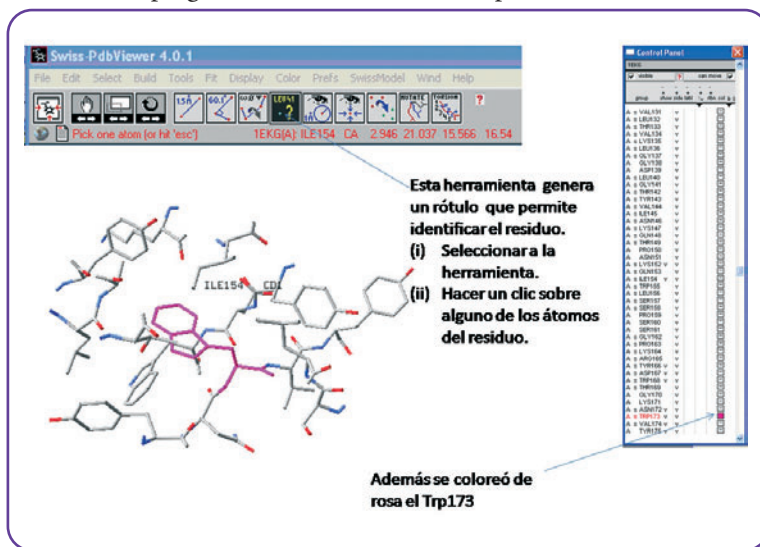
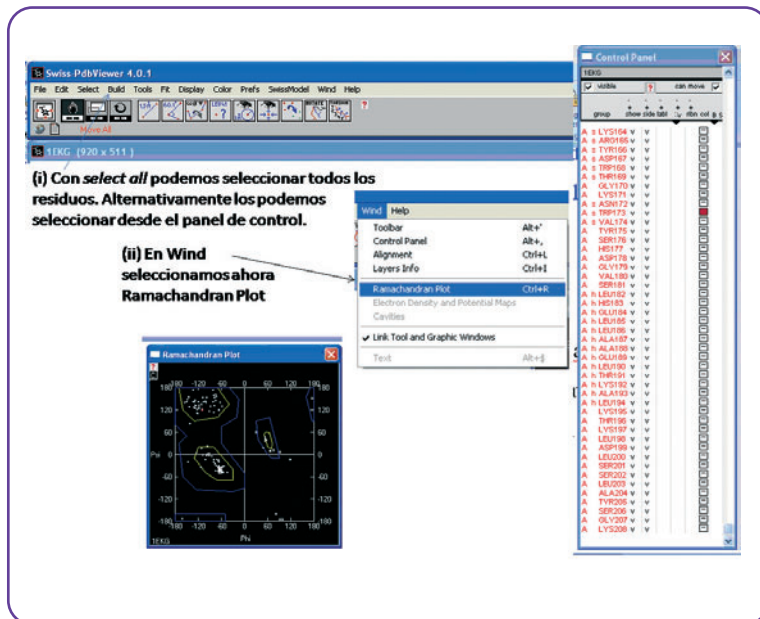


Figura P7- 2. Identificación de los residuos del entorno sobre la estructura proteica.

Respuesta P8.

Con el programa *SwissPdbViewer* el gráfico de Ramachandran puede hacerse para un grupo o para todos los aminoácidos de la proteína. Primero debemos seleccionar aquellos residuos de interés. Para esto podemos usar el panel de control o alternativamente como se muestra en la Figura podemos seleccionar todos los residuos con el comando *Select all*.

Figura P8. Gráfico de Ramachandran. En el panel de control puede verse



que se han seleccionado todos los aminoácidos (quedan en rojo).

Pueden intentar seleccionar del gráfico de Ramachandran alguno de los residuos y modificar phi o psi, moviéndose en una u otra coordenada manteniendo el botón izquierdo del mouse presionado. Es interesante verificar que le ocurre a la estructura terciaria de la macromolécula cuando modificamos tan solo un par de ángulos diedros psi y phi.

Respuesta P9.

El modelo de estado desplegado se puede construir con el programa *SPDBviewer*. Se deben seleccionar todos los residuos en el panel de control (*control panel*) y con el comando *Set omega/phi/psi* en herramientas (*Tools*) seleccionamos strand (hoja beta en inglés).

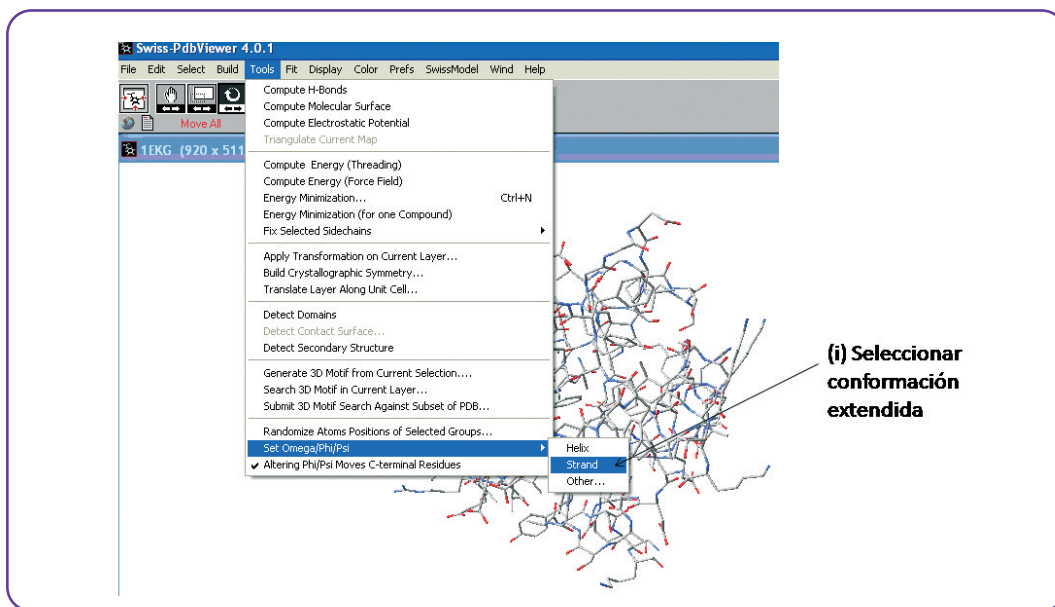


Figura P9. Preparando un modelo de estado desplegado con el programa de visualización.

Así obtendremos un modelo de estado desplegado en el que la exposición al solvente (agua) de cada residuo es máxima. Es un modelo en el que la cadena está completamente extendida. Debe observarse que este modelo es rígido y por lo tanto es un modelo poco realista ya que el estado desplegado se caracteriza por poseer flexibilidad extrema. Energéticamente será muy poco probable esta estructura.

Respuesta P10.

Para evaluar si un ángulo ω se encuentra en la configuración cis o trans es conveniente dejar visibles únicamente los residuos que forman parte de este enlace peptídico. En el caso del problema los residuos de interés son Ser158 y Pro159 de la frataxina humana. Esto se hace desde el panel de control, tildando únicamente los residuos de interés. Para poder visualizar más cómodamente al enlace será conveniente centrar el pequeño fragmento. El enlace se encuentra en la configuración trans, el ángulo diedro es de 180° . Un semiplano

queda definido por los átomos CA del residuo de serina, C (carbono carbonílico) y N, y el segundo semiplano definido por los átomos C N-CA del residuo de prolina y efectivamente la intersección entre ambos semiplanos es la recta que pasa por el enlace carbono carbonílico-nitrógeno. (C-N).

Figura P10. Visualización de un enlace peptídico en configuración trans.

Respuesta P11.

Mediante la herramienta *Compute Molecular Surface* podemos calcular el área accesible al solvente. Una esfera de 1.4 Å que simula a una molécula de agua se hace rotar sobre la superficie macromolecular. Esta esfera se pondrá en contacto únicamente con aquellos átomos que se dispongan en la superficie, los átomos del corazón de la proteína serán esencialmente inaccesibles.

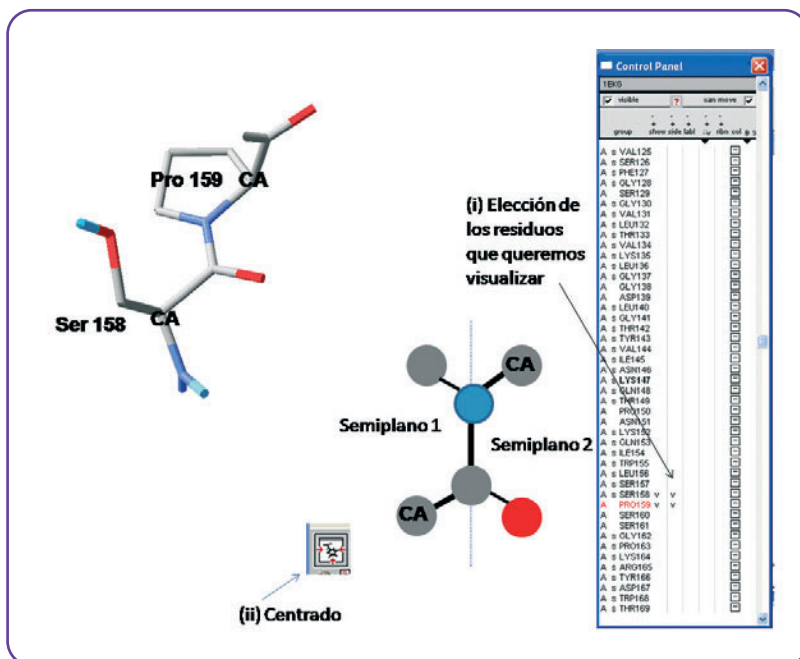
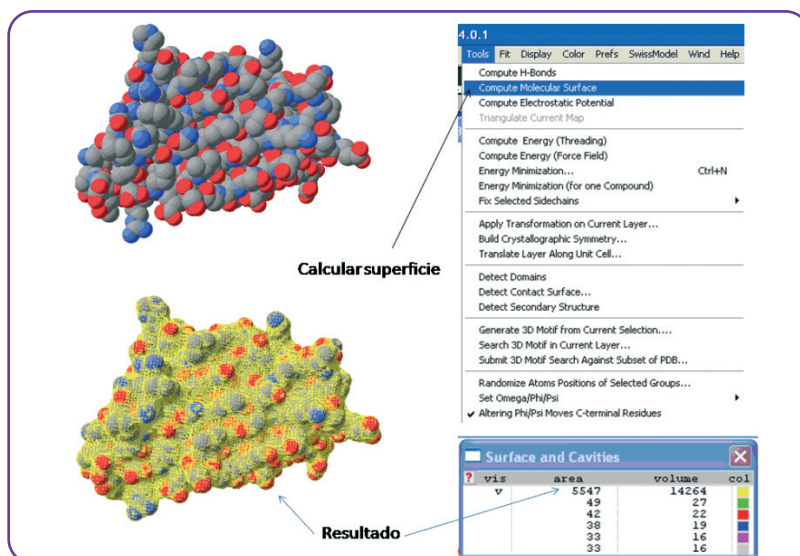


Figura P11.

Localización de la herramienta de cómputo de superficie accesible al solvente.

Este cálculo también se puede hacer sobre el modelo de estado desplegado generado en el problema 8. Con ambos datos podemos calcular la diferencia de superficie accesible entre ambos estados conformacionales. Capítulo II.



A) Para calcular la Cm (concentración media de desnaturizante) debemos tener en cuenta que en la concentración de desnaturizante $[D] = C_m$, el valor de ΔG°_{UN} es cero, esto quiere decir que el proceso es igualmente probable en cualquiera de las dos direcciones (reacción $N \rightarrow U$ y reacción $U \rightarrow N$).

$$\Delta G^\circ_{NU} = \Delta G^\circ_{H_2O_{NU}} - m_{NU} [D]$$

$$0 = \Delta G^\circ_{H_2O_{NU}} - m_{NU} [D]$$

$$\frac{\Delta G^\circ_{H_2O_{NU}}}{m_{NU}} = [D] = C_m$$

$$C_m_{\text{salvaje}} = 2.7 \text{ M}$$

$$C_m \text{ L107A} = 1.9 \text{ M}$$

B) La variante más estable es la tiorredoxina salvaje. Esto se deduce de la diferencia de energía libre de desplegado. La diferencia en Cm también es indicadora de esta diferencia de estabildades. Para la variante salvaje se requiere una mayor concentración de desnaturizante para desplegar el 50% de las moléculas.

C) Las velocidades de plegado y desplegado en la concentración media de agente desnaturizante son idénticas. Esto se deduce del hecho de que la probabilidad de plegado es igual a la de desplegado, dicho con otras palabras, en esta concentración de desnaturizante la energía del estado nativo es idéntica a la del estado desplegado.

Respuesta P13

a) $A_{280\text{nm}} = \epsilon_{280\text{nm}} \times C \times l$
 $A_{280\text{nm}} = 5690 \text{ M}^{-1} \text{ cm}^{-1} \times 35 \times 10^{-6} \text{ M} \times 1 \text{ cm}$
 $A_{280\text{nm}} = 0,199$

b) Si asumimos que por ser tan pequeños estos péptidos no poseen estructura capaz de modificar sus propiedades de absorción UV, y que por lo tanto los residuos se encuentran completamente expuestos al solvente entonces podemos usar compuestos modelos de los aminoácidos libres para calcular los $\epsilon_{280\text{nm}}$. Nuestro problema se reduce a encontrar en la bibliografía los coeficientes de extinción ($\epsilon_{280\text{nm}}$) de fenilalanina y de triptofano y tirosina.

Compuesto	Coefficientes de extinción a 280 nm ($\text{M}^{-1} \text{ cm}^{-1}$)
Triptofano	5690
Tirosina	1280
Fenilalanina	~ 0

Gill, S.C., von Hippel, P.H., "Calculation of Protein extinction coefficients from Amino Acid Sequence Data", Anal. Biochem, 182, 319-326, (1989)



Con estos coeficientes de extinción ahora podemos calcular los coeficientes de los péptidos simplemente como la suma de los coeficientes de extinción de los aminoácidos que los constituyen. AFKSTEW cuenta con un triptofano y una fenilalanina como únicos residuos aromáticos que son los que absorben en esta región espectral (240-340 nm) Pero las fenilalaninas no absorben a 280 nm, por esto

$$\begin{aligned}\epsilon_{\text{AFKSTEW}} \text{ a } 280\text{nm} &= \epsilon_{\text{W } 280\text{ nm}} = 5690 \text{ M}^{-1} \text{ cm}^{-1} \\ A_{\text{AFKSTEW}} \text{ a } 280\text{nm} &= 5690 \text{ M}^{-1} \text{ cm}^{-1} \times 35 \times 10^{-6} \text{ M} \times 1 \text{ cm} = \mathbf{0,199} \\ \epsilon_{\text{AAKSTEYY}} \text{ a } 280\text{nm} &= 2 \times \epsilon_{\text{Y } 280\text{ nm}} = 2560 \text{ M}^{-1} \text{ cm}^{-1} \times 35 \times 10^{-6} \text{ M} \times 1 \text{ cm} = \mathbf{0,09}\end{aligned}$$

Respuesta P14.

(III) Para determinar la concentración molar (M) usamos $A_{280\text{nm}} = \epsilon_{280\text{nm}} \times C \times l$. Así,
 $0.1 = 24750 \text{ M}^{-1} \text{ cm}^{-1} \times 1 \text{ cm} \times C$
 $= 4.04 \times 10^{-6} \text{ M}$

(IV) En 1L de una solución hipotética de concentración 1.0 M hay habría 1 mol de proteína disuelta o 29500g; en 1L de una solución cuya concentración es $4.04 \times 10^{-6} \text{ M}$ hay 4.04×10^{-6} moles disueltos entonces

$$\begin{aligned}1 \text{ mol} &\text{-----} 29500\text{g} \text{----} 1\text{L} \\ 4.04 \times 10^{-6} \text{ moles} &\text{---} 0.119\text{g} \text{----} 1\text{L} \\ &\mathbf{0.119\text{mg} \text{---} 1\text{mL}}\end{aligned}$$

(b) Un litro de una solución de concentración de proteína 1.0 M contiene 6.02×10^{23} moléculas de proteína (el número de Avogadro) entonces

$$\begin{aligned}1 \text{ mol} &\text{-----} 6.02 \times 10^{23} \text{ moléculas} \text{-----} 1\text{L de solución} \\ 4.04 \times 10^{-6} \text{ moles} &\text{---} 2.43 \times 10^{18} \text{ moléculas} \text{-----} 1000 \text{ mL de solución} \\ &\mathbf{2.43 \times 10^{15} \text{ moléculas} \text{-----} 1\text{mL}}\end{aligned}$$

Respuesta P15.

El triptofano libre en solución posee un espectro de fluorescencia con un $\lambda_{\text{MAX}} = 356 \text{ nm}$, mientras que en el contexto proteico del estado nativo, el λ_{MAX} puede ser menor, por ejemplo 340 nm, reflejando emisión desde un entorno apolar. El estado excitado del triptofano expuesto al solvente se estabiliza por efecto de la reorientación de los dipolos de las moléculas de agua. Esto disminuye su energía y en consecuencia la longitud de onda asociada a la transición es mayor, menos energética (ver Figura 4-21).

Respuesta P16.

A través de experimentos de dicroísmo circular podemos analizar el contenido de estructura secundaria en la región del espectro que va entre 180-250nm y el entorno de los residuos aromáticos (240-340nm), en este último caso podremos inferir la presencia de estructura terciaria.

Respuesta P17.

Las relaciones m/z para el ión más abundante y el de su izquierda son:

$$\frac{m}{z} = 1000,543 = \frac{m_p + (n)}{(n)}$$

$$\frac{m}{z} = 938,082 = \frac{m_p + (n + 1)}{(n + 1)}$$

De la primera ecuación podemos despejar m_p , la masa de la proteína.

$$1000,543 \times n - n = m_p$$

Ahora podemos introducir m_p en la segunda ecuación,

$$938,082 = \frac{1000,543 \times n - n + (n + 1)}{(n + 1)}$$

Y entonces podemos despejar n , el número de cargas.

$$938,082 \times (n + 1) - 1000,543 \times n = 1$$

$$938,082 \times n + 938,082 - 1000,543 \times n = 1$$

$$- 62.461 \times n = - 937.082$$

$$n = \frac{- 937.082}{- 62.461} = 15$$

Hemos calculado la carga relativa del ión más abundante. Ahora podemos calcular su masa:

$$\frac{m}{z} = 1000,543 = \frac{m_p + 15}{15}$$

$$m_p = 1000,543 \times 15 - 15 = 14.993,145$$

Podemos incluir el resto de los iones moleculares en el cálculo. Cada uno de ellos podrá ingresar como una nueva ecuación y al final tendremos una serie de ecuaciones con tan solo dos incógnitas. Como se usan un número grande de picos, la familia completa de iones, el cálculo es muy preciso. El espectrómetro debe estar bien calculado para que el cálculo sea exacto.

Respuesta 18.

a) Para proteínas globulares se puede calcular el radio hidrodinámico (R_S , el radio asumiendo que la macromolécula es esférica) como

$$\text{Log } R_S = 0,369 \times \text{log (PM)} - 0,254$$

Así, una proteína globular de 10063.5 Da de peso molecular debería tener un radio

$$R_S = 16.71 \text{ \AA}$$

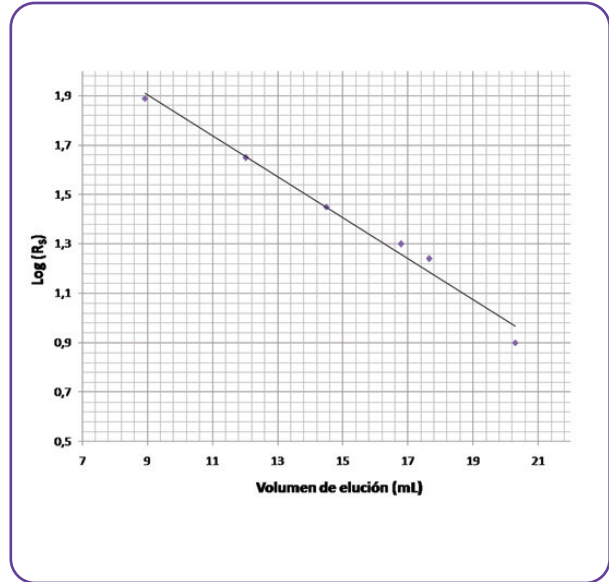
Con esta ecuación también se obtienen los radios hidrodinámicos (R_S) de las proteínas estándares de tamaño que se presentan en la tabla.

b) Grafiquemos $\text{Log}(R_S)$ en función del volumen de elución obtenido experimentalmente (ver Tabla del problema) para las proteínas estándares. Mediante una regresión lineal obtendremos luego la dependencia lineal del volumen de elución con respecto al $\text{log } R_S$.

$$\text{log}(R_S) = -0,083V_e + 2,652$$

Donde V_e es el volumen de elución. Con esta ecuación podemos calcular el volumen de elución esperado para una proteína globular monomérica de 10063.5 Da de peso molecular usando el valor de R_S obtenido ($R_S = 16.71 \text{ \AA}$).

V_e esperado para un monómero de 10063.5 Da, $V_e = 17.21 \text{ mL}$.



También podemos ingresar el valor del **volumen de elución experimental** obtenido para la proteína, 13.76 mL y podemos entonces calcular el R_S experimental que corresponde a este volumen de elución.

$$R_S = 32.37 \text{ \AA}$$

Claramente la proteína no es monomérica porque tiene un radio hidrodinámico mucho mayor que el compatible con un monómero nativo de 10063.5 Da. Si asumimos que se trata de un arreglo cuaternario globular entonces podemos calcular el peso molecular que debería tener la proteína para rendir un $R_S = 32.37 \text{ \AA}$

$$\text{Log } R_S = 0,369 \times \text{log}(\text{PM}) - 0,254$$

$$\text{Log } 32,37 = 0,369 \times \text{log}(\text{PM}) - 0,254$$

$\text{PM} = 60377.4 \text{ Da}$ que es unas 6 veces mayor que el peso molecular del monómero. Se trataría de un hexámero.

Respuesta P19.

El reemplazo por triptofano permite estudiar la accesibilidad al solvente de cada una de estas posiciones de la cadena polipeptídica siempre y cuando podamos garantizar que la mutación no produce alteraciones conformacionales. El agregado de acrilamida causará un apagamiento de la fluorescencia de triptofano por esto a concentraciones altas de acrilamida $F \ll F_0$ y el cociente $F_0/F > 1$. Cuanto más expuesto sea el triptofano más elevado será el valor de la pendiente de la recta. Por esto podemos decir que F80 posee un grado de exposición al solvente mayor.

Respuesta P20.

En ambos casos los espectros que corresponden al estado nativo son los que se grafican en línea continua. Cuando los triptofanos emiten desde un entorno apolar lo hacen con una longitud de onda mayor. Por otro lado, la disminución de la intensidad de fluorescencia en el caso A se debe probablemente a un aumento en el apagamiento por colisiones con moléculas de agua característico del estado desplegado. En el caso B el estado desplegado muestra un espectro con mayor intensidad de fluorescencia que el correspondiente al estado nativo. Esto se debe, muy probablemente, a que en ciertas proteínas existen grupos funcionales que apagan la fluorescencia en posiciones muy cercanas a los residuos que emiten. Con el desplegado, si bien aumentan las colisiones con moléculas de agua, los residuos emisores se separan espacialmente de estos grupos internos apagadores de la fluorescencia. Entre estos grupos podemos citar a los puentes disulfuro y a las histidinas. El resultado neto en este caso es un incremento de la intensidad en la emisión.

Glosario

α aminoácido un compuesto orgánico que contiene un grupo amino, un grupo carboxilato, una cadena lateral y un átomo de hidrógeno unidos a un átomo de carbono central (el carbono α).

A Amiloide acumulación de proteínas en forma de agregados insolubles, muchas veces ordenados en forma de fibrillas con alto contenido de estructura secundaria de tipo láminas β .

Anfipaticidad la propiedad de una molécula de poseer una parte polar y una apolar segregadas espacialmente una de otra.

Anfolito molécula que contiene un grupo ácido y un grupo básico.

Ångström (Å) es una unidad de longitud empleada muy comúnmente para expresar distancias interatómicas o intramoleculares. Se representa por la letra sueca Å. 1 Å escrito en notación científica es 1×10^{-10} m. Más comúnmente lo podemos escribir como 0.000000001 m.

Ángulo diedro o torsional es el ángulo entre dos semiplanos a través de una recta que es la resultante de la intersección de ambos semiplanos; un ángulo diedro queda definido por cuatro átomos sucesivos, por ejemplo el ángulo diedro ω (omega) está definido por $C\alpha_{(i)}-C'-N-C\alpha_{(i+1)}$.

Átomo pesado los átomos que no son átomos de hidrógeno, estos son: nitrógeno, carbono, oxígeno y azufre, N, C, O y S respectivamente.

B Buffer amortiguador de pH, actúa como amortiguador de pH debido a su carácter anfótero, es decir, pueden comportarse como ácidos o como bases, aceptando o donando electrones respectivamente.

C Cadena lateral es un grupo químico que está unido al carbono α y que diferencia a los aminoácidos entre sí; en el caso de la glicina es $-H$ y en el caso de la alanita es $-CH_3$.

Cadena principal el esqueleto de la proteína formado por los átomos N-C α -C de cada residuo de aminoácido. No forman parte de la cadena principal los átomos de las cadenas laterales de los aminoácidos.

Caótopo una sustancia que incrementa el desorden. El término lo usamos para describir sustancias que provocan el desplegamiento de proteínas: urea y cloruro de guanidinio solubilizando a las cadenas laterales hidrofóbicas.

Catalizador sustancia que acelera un proceso químico.

Cavidad espacio vacío en el interior de la proteína.

cDNA es el DNA copia producto de una reacción de retro transcripción a partir de mRNA

Centrifugación un método que permite separar partículas presentes en líquidos mediante la aplicación de aceleración centrífuga. Las partículas experimentan una fuerte aceleración que las obliga a sedimentar.

Código genético establece la correspondencia entre codones a nivel del DNA (y del mRNA) y aminoácidos. Cada codón de la región codificante de nucleótidos de un mRNA tiene su significado: información para un aminoácido o para terminar la síntesis de la proteína. Debemos tener en cuenta que la decodificación se realiza a través de macromoléculas que interaccionan entre sí reconociéndose.

Codón secuencia de tres nucleótidos presente en el mRNA que codifica para un aminoácido, por ejemplo el codón x codifica para el aminoácido metionina.

Concentración es la relación que hay entre la cantidad de soluto y la cantidad de disolvente en una solución, donde el soluto es la sustancia que se disuelve, el disolvente la sustancia que disuelve al soluto para formar una solución homogénea. Una menor proporción de soluto disuelto en el disolvente, implica una menor concentración de este en la disolución.

Concentración molar la relación de moles de soluto por litro de solución en una solución homogénea.

Corazón hidrofóbico también *core* o corazón apolar, grupo de residuos apolares empaquetados en el interior de las proteínas globulares.

Cromatografía de exclusión molecular es una técnica cromatográfica ampliamente difundida que permite separar macromoléculas en función de su tamaño.

Cromatografía un conjunto enorme de técnicas fundamentadas en el principio de retención selectiva. Su objetivo es separar los distintos componentes de una mezcla.

Cromóforos son los componentes de las macromoléculas que absorben luz a determinadas longitudes de onda. En las proteínas en el rango de longitudes de onda que va entre 240 y 340 nm (UV cercano) los cromóforos son los residuos de W, F y Y. Además existe una contribución de los puentes disulfuro, que también absorben en esta región espectral.

Cuerpos de inclusión proteína depositada en forma insoluble durante la expresión proteica en sistemas bacterianos.

D Da, el Dalton es la unidad de masa atómica; se la llama así en honor al químico inglés John

Dalton. Aunque también se la denomina amu (*Atomic Mass Unit*). Equivale a una doceava parte de la masa del núcleo del isótopo más abundante del carbono, el carbono-12. Es aproximadamente la masa de un átomo de hidrógeno.

Desnaturalización es la pérdida de estructura secundaria y terciaria por parte de una proteína llevándola a una conformación generalmente desordenada y biológicamente no funcional.

Dicroísmo circular absorción de luz circularmente polarizada.

Dieléctrico un material que no conduce la electricidad eficientemente, si la conduce muy ineficientemente entonces puede ser utilizado como aislante eléctrico.

Dímero se refiere a un arreglo de estructura cuaternaria en el que dos subunidades coalescen para formar un complejo macromolecular de mayor tamaño. Si ambas subunidades son idénticas lo llamaremos homodímero y si no lo son lo llamaremos heterodímero.

Dipéptido una molécula formada por dos aminoácidos unidos entre si por un enlace peptídico.

DNA ácido desoxirribonucleico (ADN) es el material genético de casi todos los organismos conocidos con la excepción ciertos tipos de virus cuyo material genético es el RNA.

DNA polimerasa enzima formada por un complejo de proteínas que cataliza la síntesis del DNA.

Dogma central de la Biología Molecular propone que existe una unidireccionalidad en la expresión de la información genética contenida en la célula. El DNA es transcrito a mRNA y éste es traducido a proteínas que efectúan la acción celular. El dogma también postula que sólo el DNA puede replicarse y transmitir la información genética a la descendencia. Fue propuesto por Francis Crick en 1970.

Dominio región de una cadena polipeptídica compacta y autoorganizada. Una unidad de plegado autónomo que posee estructura secundaria y terciaria.

E Electroforesis técnica para la separación de moléculas según la movilidad de estas en un campo eléctrico aplicado.

Electronegatividad una medida de fuerza de atracción que ejerce un átomo sobre los electrones de otro átomo en un enlace químico.

Elución es la remoción mediante la aplicación de una fase móvil adecuada de una molécula o macromolécula de una matriz cromatográfica.

Enlace covalente el enlace químico entre los átomos no metálicos de una molécula orgánica. Los electrones del enlace son compartidos por ambos átomos en un nuevo tipo de orbital denominado orbital molecular.

Enlace peptídico enlace amida que une a dos aminoácidos sucesivos en un péptido.

Enzima de restricción enzimas que catalizan el corte de la doble cadena de DNA en sitios extremadamente específicos.

Enzimas macromoléculas que catalizan reacciones químicas actuando sobre moléculas denominadas sustratos, que se transforman en los productos. Casi todos los procesos en las células necesitan enzimas para que ocurran a velocidades biológicamente compatibles.

Espectrometría de masa una herramienta para determinar la masa de una macromolécula con exactitud y gran precisión.

Estado de transición las reacciones químicas proceden a través de un estado de alta energía que tiene algunas propiedades de los reactivos y otras que ya corresponden a los productos.

Estado desplegado estado conformacional caracterizado por la ausencia de interacciones estables en el tiempo. Es un conjunto de conformaciones.

Estructura cuaternaria el nivel de estructura que resulta de la asociación de dos o más subunidades.

Estructura nativa es el estado conformacional biológicamente activo, por lo general, con estructura secundaria y terciaria bien definida. Sin embargo se debe tener en cuenta que existe una gran variedad de proteínas con estructura nativa desorganizada.

Estructura primaria está establecida por la secuencia de aminoácidos de la proteína.

Estructura secundaria estructura local que involucra aminoácidos cercanos en secuencia; los distintos tipos de estructura se definen como combinaciones periódicas de ángulos diedros phi y psi, hélices hebras.

Estructura supersecundaria en las proteínas encontramos motivos formados por la asociación y la interacción de elementos de estructura secundaria, son patrones que se repiten con regularidad en las proteínas.

Estructura terciaria arreglo espacial de elementos de estructura secundaria, empaquetamiento de cadenas laterales y el conjunto de interacciones no covalentes que determinan la estructura tridimensional de la proteína.

Eucariotas organismos que forman el dominio *Eukarya*. Estos organismos están formados por una o más células que contienen núcleo. Dentro del núcleo está encerrado el DNA nuclear. Existen muchas características increíblemente interesantes tanto funcionales como estructurales que diferencian a estos organismos de los procariotas. Entre estos últimos podemos citar a las bacterias.

Exón son las regiones de un **gen** que por lo general no son eliminadas en el proceso de maduración del mRNA. Cada exón codifica una porción específica de la proteína completa, de manera que el conjunto de exones forma la *región codificante* del gen. En el procesamiento del RNA suelen retenerse todos los exones si bien en algunos casos de *splicing* alternativo alguno de ellos puede ser eliminado con los intrones.

Fase móvil solvente que se emplea para favorecer la unión y/o elución de una molécula de una matriz cromatográfica.

F **Fluorescencia** el regreso al estado electrónico basal desde el estado excitado con emisión de luz.

Fluoróforos son los componentes de las moléculas que pueden des excitarse por vías radiativas, emitiendo luz. En las proteínas el fluoróforo intrínseco por excelencia es el residuo de W porque es extremadamente sensible al entorno químico, a la polaridad del medio.

G **Gráfico de Ramachandran** dada una estructura tridimensional proteica se grafican las combinaciones de ángulos diedros (torsionales) phi (ϕ) y psi (ψ) para cada aminoácido y puede verse fácilmente si algún aminoácido de la proteína posee una combinación no permitida de ángulos torsionales.

Grupo prostético un metal, una molécula orgánica pequeña, o una combinación de ambos que interviene de alguna manera en la función de la proteína puede estar unido tanto covalentemente como en forma no covalente a la macromolécula.

H **Hélice α** elemento de estructura secundaria que se caracteriza por 3.6 residuos por vuelta de hélice en la que los puentes de hidrógeno se hallan separados por 11 átomos de la cadena principal.

Hélice $3d_{10}$ una forma helicoidal alternativa que tiene 3 residuos de aminoácidos por vuelta y los puentes de hidrógeno separados por 10 átomos de la cadena principal.

Hidrofílico la habilidad de un átomo o de una molécula o de parte de ella para interactuar atractivamente con moléculas de agua. Los compuestos hidrofílicos son solubles en agua.

Hidrofóbica sustancia no polar con baja o nula solubilidad en agua.

Hidrólisis es la ruptura de un enlace químico por reacción con una molécula de agua. Involucra a los iones hidronio H_3O^+ o hidroxilo, OH^- , estos grupos surgen como resultado de esta transferencia de protones entre moléculas de agua cercanas.

I ***In silico*** que ocurre por simulación en una computadora.

In vitro que ocurre dentro de un tubo de ensayo, en el laboratorio.

In vivo que ocurre dentro de un organismo vivo.

Interacciones no covalentes están basadas en la existencia de fuerzas atractivas y repulsivas como puentes de hidrógeno, interacciones dipolo-dipolo, fuerzas de London. Son de origen electrostático.

Intrones regiones no codificantes del RNA que son eliminadas en el núcleo luego de la transcripción mediante un proceso conocido con el nombre de *splicing*.

L Lámina β arreglo de hebras β extendidas formando superficies estabilizadas por puentes de hidrógeno entre hebras de la cadena principal. Las hebras pueden ser paralelas o antiparalelas. Las cadenas laterales se proyectan una a 180° de la siguiente.

M Mol es la unidad con que se mide la cantidad de una sustancia, un mol de moléculas equivale a equivale a $6,02214179 \times 10^{23}$.

Molten globule un estado conformacional en el que la proteína carece interacciones terciarias fijas pero en el que el recorrido de la cadena polipeptídica en el espacio puede asemejarse al del estado nativo.

Mutación puntual cambio en la secuencia de DNA que determina el reemplazo de un aminoácido por otro. Puede ser conservativa si ambos aminoácidos poseen propiedades fisicoquímicas semejantes y no conservativas si son radicalmente diferentes.

Mutagénesis sitio dirigido técnica de biología molecular que permite alterando la secuencia de DNA de un gen reemplazar un codón por otro con el consiguiente cambio de un aminoácido por otro una vez expresada la proteína.

N NMR resonancia magnética nuclear, un tipo de espectroscopía que permite establecer la estructura tridimensional de proteínas en solución, pero que además permite estudiar los procesos dinámicos asociados con estas y otras macromoléculas.

Nomenclatura PDB es la nomenclatura utilizada para la identificación de cada átomo en un archivo de la PDB.

O Operador es una región del DNA cercana al sitio de iniciación de la transcripción que une moléculas llamadas represores que inhiben la transcripción del mRNA.

Orbitales moleculares cuando dos átomos forman una molécula, los orbitales atómicos se combinan dando lugar a nuevos orbitales (moleculares). Los orbitales moleculares π se forman como combinación de dos orbitales atómicos p por una superposición lateral. Las uniones que consisten en orbitales π son más débiles que las uniones sigma, las uniones π entre átomos ocurren sólo cuando una unión sigma se establece primero dando lugar a dobles o triples enlaces. Formalmente son funciones matemáticas que describen los estados que pueden tener los electrones en las moléculas y expresan qué probabilidad existe de encontrar electrones en determinada región del espacio.

P **PAGE** electroforesis en geles de poliacrilamida. Se trata de una técnica sencilla que permite separar proteínas sometidas a un campo eléctrico.

Paradoja de Levinthal si la búsqueda conformacional desde el estado desplegado fuera un proceso aleatorio, las moléculas deberían cambiar su conformación tantas veces que sería un proceso irrealizable en tiempos razonables. Por ejemplo, si una cadena de cien residuos tiene 10.100 posibles conformaciones (un promedio de diez conformaciones por residuo), y si una conformación se convierte en otra en el tiempo más corto posible (tal vez 10-13 ps), el tiempo promedio requerido para explorar todas las posibles conformaciones hasta hallar el mínimo energético sería 1077 años. Las observaciones, por el contrario, indican que las proteínas se pliegan en tiempos muy cortos, desde unos pocos microsegundos a minutos u horas.

pb par de bases, suele usarse como medida de la longitud de un fragmento de DNA, por ejemplo un fragmento de 1.000 pb o 1kb es un fragmento que posee una longitud de 1.000 nucleótidos; usamos indistintamente los términos base (nitrogenada) y nucleótido.

PCR Reacción en cadena de la polimerasa, amplificación (producción) de DNA en forma cíclica basada en la utilización de DNA polimerasas termo resistentes y oligonucleótidos específicos iniciadores de la síntesis del DNA.

PDB La base de datos de estructuras de proteínas, *Protein Data Bank*.

Péptido señal las proteínas suelen contener secuencias que determinarán su ubicación en los compartimentos celulares a modo de etiquetas. Estas etiquetas, en algunos casos, pueden interferir el proceso de plegado.

Péptido varios aminoácidos unidos a través de enlaces peptídicos, por lo general no se usa este término para macromoléculas de más de 40 aminoácidos de largo.

pH es la operación logaritmo anteponiendo el signo negativo y aplicada a la concentración de protones en una solución acuosa. El pH es un parámetro fundamental en biología. Un cambio en el pH puede alterar profundamente mecanismos de reacción enzimática. En los organismos vivos, el pH debe controlarse muy eficientemente.

Plásmido una molécula de DNA doble cadena circular que consta de un conjunto pequeño de genes propios, tienen la capacidad de replicarse dentro de la bacteria usando la maquinaria de replicación de DNA de la bacteria y pueden conferirle a la bacteria resistencia a algún antibiótico.

Polarizabilidad electrónica es el desplazamiento de cargas en presencia de un campo eléctrico externo, es decir en un átomo neutro la nube electrónica se reorienta de tal manera que el átomo se distorsiona y pierde su simetría.

Polipéptido una cadena de aminoácidos unidos por enlaces peptídicos.

Post-traduccional que ocurre con posterioridad a la síntesis proteica.

Procariota por ejemplo las bacterias, organismos que no tienen núcleo celular. Además existe un gigantesco número de propiedades estructurales, metabólicas y funcionales que permiten diferenciarlos de los organismos eucariotas.

Proteasa enzima que cataliza la hidrólisis de enlaces peptídicos produciendo la fragmentación de la proteína. Existe una gran variedad de proteasas, algunas de ellas son exquisitamente específicas: reconocen únicamente determinadas secuencias de aminoácidos y pueden usarse como herramientas moleculares.

Proteínas fibrosas por lo general proporcionan soporte mecánico a otras moléculas, a organelas, a células e incluso a organismos. Suelen ser insolubles y están formadas por una unidad repetitiva simple que se ensambla para formar fibras.

Proteínas globulares aquellas proteínas en las que ninguna de las dimensiones es desproporcionadamente mayor que otra. La estructura terciaria de la proteína determina su esfericidad. Los grupos apolares en su mayoría se agrupan en el interior de la molécula formando un corazón o core, mientras que los polares en su mayor parte se dispondrán en la superficie, en contacto con el solvente.

Puente de hidrógeno interacción atractiva entre el átomo de hidrógeno de un grupo donador electronegativo (por ejemplo OH o NH) y un par de electrones de un grupo aceptor (por ejemplo C=O), por la atracción de las densidades de carga mencionadas. Estos enlaces son más débiles que un enlace covalente, pero lo suficientemente fuertes como, por ejemplo, para mantener las moléculas de agua unidas a temperatura ambiente en estado líquido.

Puente disulfuro enlace entre dos átomos de azufre. En este tipo de enlace covalente intervienen dos residuos de cisteína cuyos grupos tioles se oxidan para formar el puente. Cuando es intramolecular este tipo de enlace restringe enormemente la libertad conformacional de la cadena polipeptídica.

Punto isoeléctrico el pH en el cual una proteína (u otro anfótero o polianfótero) tiene carga neta cero, $Q = 0$.

Q **Quencher** apagador de fluorescencia

Quiral: centro de asimetría por ejemplo un carbono α que tiene unidos cuatro grupos sustituyentes diferentes. El carbono α de la glicina no es quiral porque tiene dos de sus sustituyentes idénticos.

R **Radio hidrodinámico** para proteínas globulares puede calcularse asumiendo la esfericidad de la proteína, determina el coeficiente de difusión trasnacional.

Redox reacciones de óxido-reducción

Renaturalización replegado de una proteína desnaturalizada a su forma nativa.

Residuos vecinos de un residuo dado son todos aquellos residuos que se ubiquen, por lo menos en parte, dentro de una esfera de radio = 4.5 Å; siendo el centro de la esfera algún átomo de nuestra cadena lateral, por ejemplo el carbono α . de un residuo dado. Es conveniente en esta definición considerar como radio una distancia de van der Waals promedio.

S Sitio activo es una región de la estructura tridimensional de una proteína con características químicas especiales que determinan la existencia de afinidad por otra u otras moléculas.

Solución ácida una solución cuyo pH es menor que pH 7.0

Solución básica una solución cuyo pH es mayor que pH 7.0

Solución mezcla homogénea, una sustancia (solute) se disuelve en otra (solvente) formando una sola fase. Los componentes pueden separarse utilizando procedimientos físicos.

Solución neutra una solución de pH 7.0

T Traducción es un conjunto de procesos que permiten producir una proteína a partir de un mRNA, decodificando la secuencia de bases del mRNA en una secuencia de aminoácidos.

Transcripción un conjunto de procesos que culmina en la síntesis de mRNA usando como molde al DNA.

Transformada de Fourier se utiliza para pasar al dominio de las frecuencias una señal para así obtener información que no es evidente en el dominio temporal.

Transición $n \rightarrow \pi^*$ transición electrónica descrita como la promoción de un electrón desde un orbital no enlazante n (pares libres o solitarios de electrones) a un orbital antienlazante π designado como orbital π^* .

Transición $\pi \rightarrow \pi^*$ es una transición electrónica descrita como la promoción de un electrón desde un orbital enlazante π a un orbital antienlazante π designado como orbital π^* .

Transiciones electrónicas la absorción de luz UV por parte de una molécula promoviendo el paso de un electrón desde un orbital molecular fundamental a un orbital excitado. La separación energética entre estos orbitales moleculares corresponde a las frecuencias y longitudes de onda del UV. Estas transiciones se estudian mediante la Espectroscopia de absorción UV.

U UV cercano región del espectro electromagnético que se extiende desde 240 a 340 nm

UV lejano región del espectro electromagnético que se extiende desde 180 a 240 nm

Abreviaturas

ANS	(1,8)-anilinoftalensulfonato
BSA	albúmina sérica bobina
CD	dicroísmo circular
C_p	capacidad calorífica a presión constante
DNA	ácido desoxirribonucleico
dNTP	desoxinucleótidos trifosfato
ϵ	coeficiente de extinción
f_N	fracción nativa
f_U	fracción desplegada
GdmCl	cloruro de guanidinio
HPLC	cromatografía líquida de alta performance
I	estado intermedio
IPTG	β -D-isopropil-tiogalactopiranosido.
K_d	constante de disociación
K	constante de equilibrio
k	constante cinética de velocidad
k_F	constante de velocidad de plegado
k_U	constante de velocidad de desplegado
$K_{U \leftrightarrow N}$	constante de equilibrio de desplegado
N	estado nativo
NATA	N-acetilriptofanoamida
NAYA	N- acetiltirosinamida
RMN	resonancia magnética nuclear
PCR	reacción en cadena de la polimerasa
PDB	<i>Protein Data Bank</i>
PM	peso molecular
Q	rendimiento cuántico
rpm	revoluciones por minuto
RNA	ácido ribonucleico
R_s	radio de Stokes
SDS	dodecil sulfato de sodio
SDS-PAGE	electroforesis en geles de poliacrilamida en presencia de SDS
SEC	cromatografía de exclusión molecular.

FPLC	cromatografía rápida líquida de proteínas y péptidos
τ	tiempo de relajación ($1/k_F$)
T_+	estado de transición (TS)
U	estado completamente desplegado
UV	ultravioleta
$\Delta G_{N \leftrightarrow U}$	variación en la energía libre para el proceso de desplegado
λ_{MAX}	máximo de longitud de onda

Residuos de aminoácidos

Ala	A	alanina
Arg	R	arginina
Asn	N	asparagina
Asp	D	ácido aspártico
Cys	C	cisteína
Gln	Q	glutamina
Glu	E	ácido glutámico
Gly	G	glicina
His	H	histidina
Ile	I	isoleucina
Leu	L	leucina
Lys	K	lisina
Met	M	metionina
Phe	F	fenilalanina
Pro	P	prolina
Ser	S	serina
Thr	T	treonina
Trp	W	triptofano
Tyr	Y	tirosina
Val	V	valina