Elucidación estructural de proteínas en solución por RMN

Oscar Millet.

Unidad de Biología estructural, CICbioGUNE, Parque Tecnológico de Vizcaya, Ed. 800, 48160 Derio, España.

e-mail: omillet@cicbiogune.es

Recibido y aceptado el 08/07/2011

Abstract

Protein structure determination is a key step for a better understanding of molecular biology at molecular level as well as in drug development. In this context, NMR constitutes a powerful technique, provided that certain conditions apply to the sample. Moreover, pulse sequence versatility allows using NMR for studying protein-ligand interactions and biomolecular dynamics. The strengths and weaknesses of the application of NMR to protein structure elucidation are briefly discussed in the present review.

Key words: *protein structure- NMR- protein-ligand interaction*

Introducción

La aplicación de la química en la biología constituyó, a principios del siglo XX, el principal detonante de la exploración de la estructura de la materia biológica. Sin embargo, hubo que esperar al desarrollo de diferentes técnicas espectroscópicas, y no fue hasta mediados de los años cincuenta que Perutz y Kendrew obtuvieron la estructura cristalográfica de la mioglobina de esperma de ballena, la primera estructura tridimensional de una proteína (1). En la misma década, Watson y Crick interpretaron correctamente los datos cristalográficos proporcionados por Rosalind Franklin y pasaron a la historia por descubrir la doble hélice del ADN (2). Desde entonces la importancia de la biología estructural no ha parado de crecer y hoy la elucidación de la arquitectura molecular es fundamental para la comprensión de los procesos biológicos así como en el diseño de fármacos. Sin embargo, a pesar del esfuerzo invertido, la estructura de biomacromoléculas sigue siendo una tarea que se puede considerar de difícil a imposible, dependiendo del sistema de estudio. Los trabajos de Ernst y Wüthrich, ambos de nacionalidad Suiza y ambos premio Nobel, han revolucionado la resonancia magnética nuclear (RMN) adecuándola para el estudio de sistemas biológicos (3, 4). Actualmente la RMN permite, en casos favorables, obtener la estructura tridimensional de proteínas, ácidos ribonucleicos y fragmentos de ADN (5, 6). En la presente actualización se discutirá el proceso de elucidación estructural por RMN, haciendo especial énfasis en los últimos desarrollos de la técnica así como en las ventajas y las limitaciones de la misma.

Conceptos básicos de RMN

En presencia de un campo magnético permanente se produce un desdoblamiento de niveles energéticos de los núcleos atómicos debido al efecto Zeeman. Las poblaciones de equilibrio de dichos niveles pueden alterarse mediante pulsos de radiofrecuencia. La resonancia magnética nuclear consiste en la medida de la señal que se produce en el retorno del sistema al equilibrio. Estas mediciones proporcionan información de la distribución de niveles cuánticos que, en grado último, son una expresión de la constitución de la materia. El observable más común en RMN se denomina desplazamiento químico que, de manera simplificada, equivale a la posición que cada señal ocupa en el espectro. El desplazamiento químico es muy importante porque permite diferenciar por el tipo de núcleo observado (carbono, protón, nitrógeno,...), por las características químicas del núcleo bajo consideración (protones alifáticos, aromáticos,...) y por el entorno local de cada núcleo (protón rodeado de protones frente a un protón cercano a carbonos). Así, el desplazamiento químico constituye un exquisito descriptor de la estructura de la materia y por ello la RMN, tras ser inventada por los físicos, ha interesado a químicos primero y después a biólogos (7).

Los átomos que están unidos por enlaces producen un desdoblamiento de señales en el espectro debido al acoplamiento escalar entre los núcleos involucrados. Este desdoblamiento puede aumentar la complejidad del espectro porque incrementa el número de señales (de hecho hay métodos para eliminar este efecto) pero proporciona una valiosa fuente de información para establecer la topología química de la molécula. El acoplamiento escalar ha permitido diseñar la espectroscopía multidimensional en la que los desplazamientos químicos de un tipo de núcleos (p. ej. protón) se comparan con los de otro núcleo (p. ej. carbono) en un espectro bidimensional en el que tan solo aparece una señal en el plano protón/nitrógeno en aquellas coordenadas que haya un par de núcleos que están acoplados escalarmente. Un segundo tipo de efecto (el acoplamiento dipolar) es proporcional al ángulo que forma el espín nuclear con el eje principal del campo magnético externo y también es susceptible de proporcionar información estructural. El problema es que, debido al movimiento de la molécula que se produce en solución, todas las orientaciones son equiprobables y la componente neta de dicho acoplamiento se promedia a cero. Hace algunos años, Tjandra y Bax utilizaron bicelas lipídicas para orientar a las moléculas en solución con lo que se reintroducía el acoplamiento dipolar sin perjuicio del resto de propiedades espectrales (8, 9). Este procedimiento se ha demostrado muy exitoso y ha conllevado la aparición de un número elevado de medios orientadores diferentes.

Quizás el observable más importante en RMN es el efecto nuclear Overhauser (el denominado NOE) que consiste en la medida de la una velocidad de relajación cruzada entre dos núcleos. Es relevante porque esta información proporciona una estimación de la distancia más corta entre los dos núcleos involucrados, independientemente de que estén unidos por enlaces o no.

RMN de proteínas en solución

Debido a que la diferencia de niveles anteriormente citada es energéticamente muy pequeña, la RMN es una técnica muy insensible. Esto implica que hay que trabajar con concentraciones elevadas de proteína y, dado que la evolución no ha diseñado a estas moléculas para tal fin, fenómenos de agregación y/o precipitación ocurren con relativa frecuencia. Las casas comerciales que fabrican los imanes de resonancia son conscientes de este problema y el estudio de biomoléculas es el motor para que estas compañías diseñen imanes y dispositivos cada vez más potentes. Un ejemplo exitoso de esta investigación lo constituyen las sondas criogénicas que aumentan mucho la sensibilidad del equipo con la subsiguiente reducción en la concentración efectiva de proteína en la muestra.

No todos los isótopos nucleares son sensibles al fenómeno de la RMN. De hecho, de los átomos más importantes que constituyen a las proteínas (hidrógeno, carbono, nitrógeno y oxígeno) tan solo el protón presenta desdoblamiento de niveles energéticos en presencia de un campo magnético y es casi 100% el constituyente de los núcleos de hidrógeno. En el caso del carbono y del nitrógeno, los isótopos activos son el C13 y N15 que presentan abundancias naturales del 1,2% y 0,1% respectivamente. Así, para "activar" estos núcleos para ser estudiados por RMN, se acostumbra a preparar muestras marcadas en las que se enriquecen en estos isótopos nucleares. Esto se puede conseguir mediante técnicas de biotecnología en las que las fuentes de carbono y nitrógeno están controladas (10). Por último, el oxígeno no presenta buenas propiedades espectroscópicas y no se suele estudiar en RMN biomolecular. Los diferentes experimentos de RMN específicamente diseñados para proteínas se han diseñado a partir de las características químicas de las mismas. El experimento más importante que se aplica es el HSQC (del inglés heteronuclear single quantum correlation). Este experimento correlaciona los núcleos de protón que se encuentran a un enlace de un núcleo de nitrógeno. En el caso de las proteínas esta situación se da una vez para cada aminoácido (en su enlace peptídico). Así, en el espectro HSQC de una proteína aparecerá una señal por cada aminoácido y este experimento constituye un verdadero "carné de identidad" de la proteína. Contando el número de señales en el espectro ya sabemos cuantos aminoácidos tiene la proteína, aunque esta sería una información que se puede obtener de manera mucho menos costosa mediante otras técnicas. Más interesante es la información que proporciona la posición de las señales en el espectro. Tal y como se muestra en la Figura 1, las proteínas que están desplegadas muestran una dispersión de señales muy pobres, sobretodo en la dimensión de protón y que contrasta con la diversidad de desplazamientos químicos que se observan en una proteína con estructura terciaria (o cuaternaria). Esto es así porque la conformación plegada añade una contribución adicional al desplazamiento químico que permite dispersar aún más las señales.

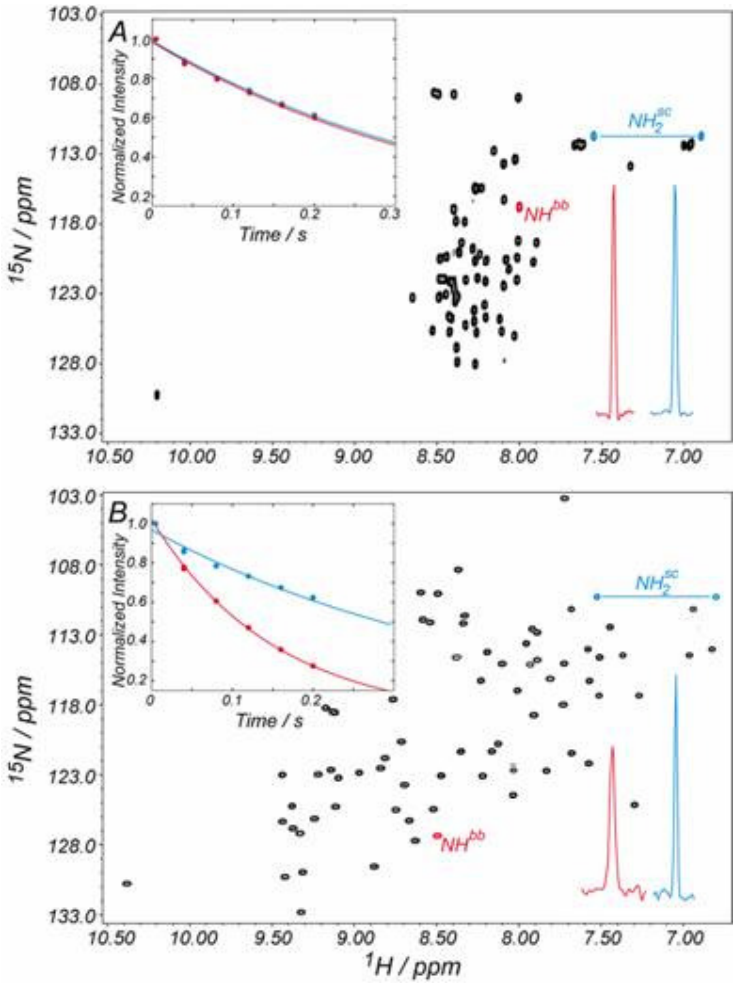

Figura 1: Espectro HSQC de una proteína en condiciones de desplegamiento (arriba) y en condiciones que favorecen el plegamiento de la misma (abajo).

Asignación espectral

Con el fin de proceder a la elucidación estructural, la primera tarea que se debe realizar es la asignación de los espectros de RMN. Dicho de otra manera, se deben identificar cada una de las señales del espectro (p. ej. el de la Figura 1) con los correspondientes aminoácidos de la secuencia de la proteína. Para ello se utiliza nuevamente el acoplamiento escalar y se registran una serie de experimentos que nos relacionan los núcleos que están conectados por enlaces. Algunos de estos experimentos nos dan información acerca de la naturaleza del residuo (p. ej. si el carbono alfa pertenece a una leucina o a una prolina). Otros experimentos nos permiten establecer la conectividad con el aminoácido precedente para así encajar las piezas (aminoácidos) en el rompecabezas (la secuencia). En principio debiera ser posible obtener la información de la asignación de la totalidad de señales en el espectro. Sin embargo, en la práctica se suele asignar entre un 85% y un 98% de las mismas, debido a problemas de solapamiento de señales y/o de desaparición de las mismas por otros fenómenos (relajación, intercambio, etc...).

El tamaño de la proteína es una variable importante a la hora de abordar la asignación espectral. Obviamente la complejidad del proceso aumenta linealmente con el número de aminoácidos implicados pero quizás más importante es que la probabilidad de que dos señales compartan desplazamiento químico (solapamiento de señales) aumenta también. Finalmente existe un tercer problema derivado del aumento del peso molecular de la molécula que es el del ensanchamiento de la señal debido a un movimiento más lento de la molécula en la solución. Como la integral de la señal depende solo del número de núcleos implicados (que no cambia), un ensanchamiento de señal inevitablemente va acompañado de una disminución de la relación señal/ruido. Este fenómeno ha limitado el tamaño máximo de moléculas que se podían utilizar hasta 15-20 KDa. El ingenioso desarrollo de un método de compensación de mecanismos de relajación (denominado TROSY) ha permitido resolver este problema en gran parte y, actualmente se puede trabajar más o menos rutinariamente con proteínas de hasta 40 kDa (11, 12). El laboratorio de Lewis Kay (Toronto) ostenta la proteína más grande jamás resuelta por RMN, la malato sintasa G de 82 kDa (13, 14).

Medida de observables

Tal y como se ha indicado anteriormente, el NOE constituye el principal observable en la determinación de la estructura de proteínas por RMN. En general se considera que todos los pares de núcleos que se encuentren a una distancia igual o inferior a 5 Å son susceptibles de producir un NOE. Dichos NOES se detectan en un experimento multidimensional denominado NOESY en el que un determinado NOE aparece como un pico de cruce situado en las coordenadas de los desplazamientos químicos de los dos núcleos implicados. En principio, la intensidad del pico es inversamente proporcional a la sexta potencia de la distancia y, de esta relación, se debieran establecer disposiciones espaciales más o menos exactas. El problema es que las proteínas no son entidades estáticas y esta distancia está en constante fluctuación. El movimiento de los núcleos contribuye con otros términos de la ecuación de una manera mucho menos predecible, haciendo que la extracción de distancias exactas a partir de los NOEs no sea posible. Sin embargo, la intensidad del NOE se utiliza como un indicador cualitativo acerca de la cercanía (o lejanía) de los dos núcleos que lo generan. El acoplamiento escalar también proporciona información estructural (además de la componente topológica ya mencionada anteriormente). El acoplamiento a tres enlaces es función del ángulo diedro que lo conforma. Los elementos de estructura secundaria adoptan posiciones específicas en el ángulo diedro (típicamente representadas en el diagrama de Ramachandran). Así, la medida de los acoplamientos escalares a tres enlaces es útil para el refinamiento de la estructura (vide infra).

Igualmente, los acoplamientos dipolares convenientemente reintroducidos con ayuda de un medio orientador proporcionan una valiosa información estructural. Este observable está especialmente indicado para obtener información a larga distancia (por ejemplo la orientación de dos dominios diferentes de una misma proteína) (15), dado que es muy complementaria a la obtenida con el NOE.

Finalmente, el propio desplazamiento químico contiene una contribución conformacional y ayuda en el proceso de refinado de la estructura. Además, cabe destacar que se pueden utilizar otros observables como las restricciones paramagnéticas que también contienen información estructural.

Calculo de la estructura

El conjunto de restricciones experimentales medidas (ver apartado anterior) se utiliza para establecer un modelo computacional que es consistente con todos los datos. Para ello se utiliza la mecánica molecular modulada en presencia de un campo de fuerzas (16). Inicialmente se parte de una conformación de la proteína totalmente extendida. El programa de modelado va probando las diferentes conformaciones hasta encontrar una que satisfaga las siguientes condiciones: i) cumpla con los preceptos del campo de fuerzas impuesto (en términos de ángulos y distancias) y ii) ponga en acuerdo al mayor número posible de restricciones experimentales. Para ello el programa explora el espacio conformacional mediante un método denominado simulated annealing que se asemeja a los ciclos de frío (poco movimiento) y calor (mucho movimiento) en el proceso de forja de un metal. Al final se obtiene una batería de estructuras que minimizan las restricciones experimentales (que contienen el menor número de violaciones) y que se considera el modelo refinado. A diferencia de otras técnicas, en RMN no se considera únicamente una estructura sino que se representan varias de ellas (ver Figura 2) con energías similares para ver la calidad de los datos experimentales: cuanto mayor sea el número de restricciones experimentales utilizadas, más parecidas van a ser estas estructuras (equivalente a mayor resolución).

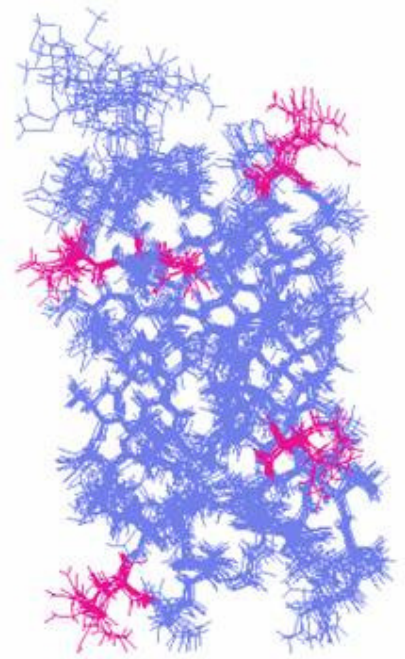

Figura 2: superposición de las diez estructuras de más baja energía para el dominio de unión a IGg de la proteína L de *streptococcus magnus*.

Consideraciones finales

Tal y como se ha descrito en la presente actualización, la RMN constituye una poderosa técnica para la elucidación estructural de proteínas siempre que la muestra cumpla una serie de propiedades y con la ventaja de que proporciona información en solución. Es de remarcar que la RMN es una técnica muy versátil y que permite realizar estudios de caracterización de la dinámica de proteínas y de sus interacciones con otros ligandos y efectores proteicos. Por todo ello, la RMN ocupa un lugar central en la biología estructural moderna.

Referencias

1. Kendrew JCPerutz MF (1957) X-ray studies of compounds of biological interest *Annu Rev Biochem* 26: 327-372.
2. Watson JDCrick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid *Nature* 171: 737-738.
3. Ernst RR (1992) Nobel Lecture. Nuclear magnetic resonance Fourier transform spectroscopy *Biosci Rep* 12: 143-187.
4. Wuthrich K (1989) Determination of three-dimensional protein structures in solution by nuclear magnetic resonance: an overview *Methods in enzymology* 177: 125-131.
5. Tugarinov V, Hwang PMKay LE (2004) Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins *Annu Rev Biochem* 73: 107-146.

6. Shajani ZVarani G (2007) NMR studies of dynamics in RNA and DNA by 13C relaxation *Biopolymers* 86: 348-359.
7. Ernst RR (1987) Methodology of magnetic resonance imaging *Quarterly reviews of biophysics* 19: 183-220.
8. Tjandra NBax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium *Science (New York, N.Y* 278: 1111-1114.
9. Tjandra N, Omichinski JG, Gronenborn AM, Clore GMBax A (1997) Use of dipolar 1H-15N and 1H-13C couplings in the structure determination of magnetically oriented macromolecules in solution *Nature structural biology* 4: 732-738.
10. Goto NKKay LE (2000) New developments in isotope labeling strategies for protein solution NMR spectroscopy *Current opinion in structural biology* 10: 585-592.
11. Pervushin K, Riek R, Wider GWuthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution *Proceedings of the National Academy of Sciences of the United States of America* 94: 12366-12371.
12. Riek R, Pervushin KWuthrich K (2000) TROSY and CRINEPT: NMR with large molecular and supramolecular structures in solution *Trends Biochem Sci* 25: 462-468.
13. Grishaev A, Tugarinov V, Kay LE, Trewhella JBax A (2008) Refined solution structure of the 82-kDa enzyme malate synthase G from joint NMR and synchrotron SAXS restraints *Journal of biomolecular NMR* 40: 95-106.
14. Tugarinov V, Choy WY, Orekhov VYKay LE (2005) Solution NMR-derived global fold of a monomeric 82-kDa enzyme *Proceedings of the National Academy of Sciences of the United States of America* 102: 622-627.
15. Millet O, Hudson RPKay LE (2003) The energetic cost of domain reorientation in maltose-binding protein as studied by NMR and fluorescence spectroscopy *Proceedings of the National Academy of Sciences of the United States of America* 100: 12700-12705.
16. Lopez-Mendez BGuntert P (2006) Automated protein structure determination from NMR spectra *Journal of the American Chemical Society* 128: 13112-13122.

QuímicaViva
ISSN 1666-7948
www.quimicaviva.qb.fcen.uba.ar

Revista ***QuímicaViva***
Número 2, año 10, Agosto 2011
quimicaviva@qb.fcen.uba.ar